



# Thinking in Words: Implicit Verbal Activation in Children and Adults

## Citation

Khan, Manizeh. 2013. Thinking in Words: Implicit Verbal Activation in Children and Adults. Doctoral dissertation, Harvard University.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11156823>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

© 2013 – Manizeh Khan  
All rights reserved.

## Thinking in words: Implicit verbal activation in children and adults

### Abstract

The relationship between language and thought has long been a topic of interest and controversy in cognitive science. In this dissertation, I address one aspect of this issue: when is language present during internal thought? Simple introspection tells us that we sometimes use inner speech, but is this the exception or the rule? Using eye-tracking measures, we investigated whether infants, children and adults implicitly activate verbal labels while silently looking at pictures of objects. In the first study, 4-year-olds, 7-year-olds and adults completed a working memory task. While the two older age groups spontaneously chose a verbal encoding strategy for the pictorial stimuli, the 4-year-olds did not, suggesting a late emergence for implicit language use. The second study, however, challenges this conclusion as we find evidence for spontaneous implicit verbal activation in 24-month-old infants during free-viewing of pictures of familiar objects. The final study provides a more detailed look at the nature of the implicit verbal representations that are activated in adults during visual image processing. Unlike the 24-month-old infants, and unlike adults engaged in a working memory task, adults in this visual image processing task did not robustly activate phonological representations but did show some evidence of lexical activation, perhaps at a more abstract level of representation. Taken together, these results suggest that: 1) even very young children spontaneously engage inner speech, 2)

adults and children use implicit verbal labeling in different ways, and 3) different tasks can evoke different levels of implicit verbal activation.

# Contents

Acknowledgements.....	v
Chapter 1. Introduction.....	1
Chapter 2. Spontaneous retrieval of object names in a short-term memory task: Changes over development.....	9
Chapter 3. Spontaneous verbal labeling in 24-month-old infants.....	31
Chapter 4. Implicit activation of verbal labels during visual object processing.....	48
Chapter 5. Conclusion.....	75
References.....	82
Appendices.....	93

## Acknowledgements

This dissertation was made possible by the support of many people, too many to mention them all in this space. First and foremost, I am deeply grateful to my dissertation advisor, Jesse Snedeker. The studies in this thesis simply would not have existed without her collaboration, advice, support and inspiration. I feel incredibly lucky that I have been able to spend these years talking to and learning from someone so insightful, knowledgeable and generous.

I owe many thanks to my other committee members, Elizabeth Spelke, Susan Carey, Felix Warneken, and George Alvarez, for their support. I am particularly grateful to Liz, who has consistently shown such enthusiasm for the work in this thesis. I always feel more excited about my projects after meeting with her, and I only wish that I had done it more.

I, of course, am deeply appreciative of the work of Whitney Fitts, Shan Wang, and Amy Geojo, my collaborators on the second and third studies in this dissertation. I am in awe at the level of intellect and sheer herculean effort both Whitney and Shan brought to these projects. Working with Amy is so much like simply talking about interesting topics with a good friend that it hardly even feels like work. I am grateful to Amy for so much more than this project. She and Joshua Hartshorne have been there every step of the way and I am so glad to have gone through graduate school with such intelligent, supportive and friendly people.

There are many other lab members, both current and former, to whom I am indebted for so many reasons. Hugh Rabagliati, Eva Wittenberg, Mahesh Srinivasan, Yi Ting Huang, Tim O'Donnell, Rebecca Nappa, Duygy Ozge and Noemi Hahn have all been the best colleagues, friends, editors, consultants, critics and my work is infinitely better for having

been surrounded by them. Hugh Rabagliati deserves particular mention for his patience as he coached me through the statistical analysis technique used in Chapters 3 and 4.

Everything is easier because of the practical support of Jesse's lab managers: Carlyn Friedberg, Kate McCurdy, Tracy Brookhyser, Margarita Zeitlin, Jean Crawford, Elizabeth Casserly, Amanda Worek. And nothing would exist without the tireless work of research assistants. Amy Chen and Edita Zhabjaku worked on these studies, and I have been fortunate to work with many other talented students on different projects.

I also have to include my gratitude to Celia Raia here. Ultimately, Celia is the one who makes everything work, who knows the answer to every question, and who helps confused students like myself figure out how to graduate.

Finally, I must acknowledge the endless support of my parents, Faiyaz and Minnat Khan, who have cheered me on from afar. And John Campbell, who makes everything more lovely.

## Chapter 1. Introduction

What is language for? The most obvious use, of course, is communication. However, it also clear that we often use language even in the absence of communication. We silently argue the pros and cons of different alternatives before making a decision. We repeat a phone number to ourselves for the duration of the short walk from the phonebook, or computer monitor, to the phone. We ask everyone else in the room to be quiet so that we can think through the numbers involved in a calculation. What is the role of language in these situations? Is it epiphenomenal or does it serve a cognitive function? What determines when we do and when we do not call up linguistic representations in our private thoughts?

Understanding the parameters of inner speech – when does it emerge in development which task contexts elicit it, and what is the nature of the representations involved? – is critical to a range of theories concerning the relationship between language and thought. Many scholars have written about the relationship between language and thought, often arriving at the conclusion that the most important functions of language are in fact separate from communication. In discussing the evolution of language, Chomsky (2007) suggests that in fact the cognitive architecture of language evolved for thought rather than communication. Berwick, Friederici, Chomsky and Bolhuis (2013) expand on this position: “One key implication is that communication, an element of externalization, is an ancillary aspect of language, not its key function, as maintained by what is perhaps a majority of scholars. Rather, language serves primarily as an internal ‘instrument of thought’” (p. 90). Studying the nature of inner speech is of direct relevance to such hypotheses. To the extent that language is an instrument of thought, and not simply a tool



for communication, we should observe the activation of linguistic representations in non-communicative contexts.

Setting aside the question of the primary function of language, the study of inner speech bears on how language interacts with nonverbal cognition. Whorf's (1956) linguistic relativity hypothesis proposes that the language one speaks has deep consequences for the thoughts one can think. Investigating the extent to which thought spontaneously involves linguistic representations provides a different angle for examining this hypothesis. Perhaps language does not change what thoughts you *can* think, but rather language is often the medium for the thoughts you do think. If the latter is true, then we might see speakers of different languages performing non-communicative cognitive tasks differently, but this would reflect the verbal representations that they are using to perform the tasks, and would not indicate underlying divergences in their conceptual representations.

The relationship between language and thought is also central in Spelke's (2003) theory of combinatorial thought, in which she discusses uniquely human cognitive capacities. Spelke suggests that language provides a powerful combinatorial tool for cognition. In particular, she proposes that it is natural language that allows humans, and not other primates, to combine information across the otherwise modular components of core cognition. Finding widespread implicit verbal activation in non-communicative contexts would add credence to the hypothesis that concepts are combined using linguistic representations, since it would demonstrate that the proposed mechanism is active during cognitive processing.

Vygotsky (1934/1987), discussed self-directed speech from a developmental perspective, and drew a distinction between audible private speech and silent inner speech. He proposed that language begins as an external social tool, which is gradually co-opted as a cognitive tool. As part of this process, children go through an intermediate stage of overt private speech in their preschool years. Private speech becomes internalized in the early elementary school years, and continues to become more abstract and dissimilar from overt speech over development. In contrast to the theories discussed above, which discuss cognitive functions as central to language, Vygotsky's approach suggests a relatively late emergence for language-in-thought. Non-communicative functions are even further removed from the development of language in Piaget's *The Language and Thought of the Child* (1923/1962), in which preschoolers' private speech is described as social speech that is poorly executed due to egocentrism, and so it is not as conceived of as a cognitive tool. On Piaget's view, language development is initially closely tied to cognitive development, as both rely on symbolic representations, but language is not actually involved in internal thought.

Across the varied perspectives outlined above, there is the common question of the relationship between language and private thought. The goal of this dissertation is to explore this relationship by asking: When is language present in silent thought? From infancy to adulthood, is thinking in words the exception or the rule? What levels of representation are accessed in inner speech - is inner speech always phonological in nature or can it exist at higher levels of linguistic representation?

Current theories and empirical findings on inner speech in adults do not paint a clear picture (for review, see Martinez-Manrique & Vicente, 2010; Vicente & Martinez-

Manrique, 2011). Estimates of the average occurrence of inner speech, based on self-reports obtained from experience-sampling, range from approximately 25% of the time (Heavey & Hurlburt, 2008) to 75% (Klinger & Cox, 1987). Descriptions of the phenomenology of inner speech are also inconsistent. Vygotsky (1934/1987) suggests that, in the adult state, inner speech is a syntactically simplified form of outer speech that is also degraded in other ways, such as pitch. Martinez-Manrique and Vicente (2010) point out that inner speech can often occur in sophisticated full sentences, although this is most apparent when mentally rehearsing for future external speech or writing. Fernyhough (2004) reconciles the existence of both impoverished and elaborate inner speech by suggesting that these phenomena occur at different levels of processing. In contrast, in discussing the role of language in consciousness, Jackendoff (1996) specifically targets phonology as the level of representation involved in conscious inner speech. To gain traction on any of these approaches, it is clear that more empirical evidence is needed on the presence and form of internal verbal representations in different contexts.

Research that addresses the development of internal verbalization also does not yield a consistent picture. Much of this work explores Vygotsky's (1934/1987) proposal that speech is initially solely a social tool and gradually becomes internalized. Studies have documented that audible self-directed speech begins around 2-years of age and becomes more covert through the preschool and early elementary school years. This trajectory is taken as evidence for the emergence of inner speech (Berk, 1986; Fernyhough, 2009; Winsler & Naglieri, 2003; see Winsler, 2009 for review). However, there is little direct evidence for the presence or absence of inner speech during these years (but see Al-Namlah, Fernyhough, & Meins, 2006).

Another line of work that speaks to the development of internal verbal representations comes from the working memory literature. While adults routinely make use of verbal representations, such as in the phonological loop (Baddeley & Hitch, 1974), to maintain information in working memory, such strategies are not observed in young children. Preschoolers do not show effects of word length, phonological similarity or verbal suppression when memorizing images, suggesting that they do not use internal verbalization in these tasks (Conrad, 1971; Ford & Silber, 1994; Hayes & Schulze, 1977; Hitch & Halliday, 1983; Hitch, Halliday, Dodd, & Littler, 1989; but see Henry, Tuner, Smith, & Leather, 2000; Hulme, Silvester, Smith, & Muir, 1986). However, it is possible that young children do activate internal verbal representations in these tasks during encoding, but do not use verbal rehearsal, so there is no trace of the verbal representation remaining at test (Hitch & Halliday, 1983; Johnston & Conning, 1990; Nairne, 1990).

While studies on the development of working memory tend to argue against implicit verbal activation in young children, recent results from infant lexical development studies suggest that implicit verbalization might in fact be present very early in childhood. In a series of priming studies, 18- and 24-month-old infants saw a prime image, presented in silence, followed by a split screen of the target and the distractor images, at which point they heard the target label (Mani, & Plunkett, 2010; 2011; Mani, Durrant, & Floccia, 2012). Infants' looking behavior is affected by the phonological relationship between the labels for the target and prime images, even though the prime image is never explicitly labeled, indicating that the infants spontaneously activated phonological representations for the images in these studies. Thus, these studies suggest that visually presenting images of objects leads to phonological activation of the object labels in 18 and 24 month old infants,

at least in a semi-verbal context. It is an open question whether implicit labeling is a spontaneous reaction to visual object encoding at this age, or if this process is elicited by the fact that other images are explicitly labeled in these studies, perhaps creating a “naming game” task context.

The current studies focus on the phenomenon of implicit verbal labeling of images as a window into the development of inner speech. There are several reasons for taking this approach. Targeting lexical activation is an attractive starting point for investigating internal linguistic representations. Gaining an understanding of implicit lexical activation will provide a solid foundation for future work examining more complex linguistic structures in which these lexical items are combined. Using single-word activation, with concrete nouns in particular, has the additional advantage of allowing comparisons across age groups. We can look for implicit verbal labeling even in infants, who are not yet producing complex linguistic structures, and use similar paradigms to probe these representations in adults.

There are also other methodological reasons for focusing on implicit verbal labeling. It is more apparent how to evoke and measure concrete noun activation compared to more complex inner speech. In order to measure inner speech, we need a paradigm where we can reasonably predict what linguistic representations would be produced if inner speech is occurring. While it is straightforward to find images that reliably elicit one particular label, such as “apple”, it is more difficult to find stimuli that reliably elicit one particular sentence.

The following chapters present three studies that explore implicit verbalization in different age groups, first in a visual working memory task and then in a less constrained

visual object processing task. In addition to the specific knowledge gained through this research, the studies presented here provide examples for how to probe the representations involved in inner speech without relying on introspection and self-report.

Chapter 2 investigates the development of internal language use by measuring looking times to objects during encoding in a short-term visual memory task, based on prior work with adults (Zelinsky & Murphy, 2000). We recorded eye movements while participants from different age groups encoded displays consisting of two images with short names and two images with long names. Longer looking times to the pictures of objects with longer names, as opposed to short names, would indicate that verbal label retrieval has occurred. Using this paradigm, we explore the development of verbal encoding of visual stimuli, comparing 4-year-olds, 7-year-olds and adults.

Chapter 3 shifts the focus to a younger age group. Curiously, given the repeated failure to find implicit verbal activation in preschoolers in the working memory literature, the lexical priming studies conducted by Mani and colleagues indicate that verbal encoding does occur in infants (Mani, et al., 2012; Mani & Plunkett, 2010; 2011). These studies seem to suggest that infants engage in implicit verbal encoding in non-communicative contexts. However, they differ from implicit verbal encoding studies with adults in that some parts of the task involve explicit verbal labeling. In these studies, there is an alternating pattern between images that are silently presented and explicitly labeled images; the infants implicitly label the former. This verbal scaffolding might induce verbal encoding in infants, so we cannot conclude that this is something infants spontaneously engage in. We present a study that explores whether implicit verbal activation in infants is an artifact of this naming

game context, or if it is in fact spontaneous, by adapting the earlier paradigms to be completely nonverbal.

Chapter 4 explores implicit verbal activation in adults in a picture viewing task that parallels the one used in the infant studies. The existing literature suggests that phonological activation occurs in infants during picture viewing but studies to date are much more equivocal with respect to implicit labeling in adults (e.g. Meyer, Belke, Humphreys, & Telling, 2007; Telling, 2008; Zelinsky & Murphy, 2000). One source of ambiguity in evaluating the differences between adults and infants, and in reconciling divergent results from various adult studies, is that these studies use a variety of different paradigms. Here we discuss two experiments that use a paradigm based on the one used with infants in Chapter 3. We included a range of different types of stimuli in this priming paradigm in order to both investigate whether implicit verbal activation occurs and also learn more about the nature of these implicit representations. Specifically, we looked at both homophone priming and phonological-onset priming and found different results for these two types of trials. The data pattern suggests that implicit lexical activation might occur in the absence of phonological activation.

Chapter 5 summarizes and synthesizes the results from Chapters 2, 3 and 4, and discusses different possible developmental trajectories. The question of whether implicit verbal encoding becomes more or less prevalent with age does not have a simple answer. Rather, the developmental theory must take into account the task context and how this might interact with age. The distinction between automatic and strategic verbal encoding is also critical in understanding the trajectory of this type of inner speech. Finally, it is also

important to consider that different levels of representation might be recruited for implicit labeling in different contexts.



## Chapter 2. Spontaneous retrieval of object names in a short term memory task: Changes over development<sup>1</sup>

The most obvious uses of language occur when information is being communicated. However, this is not the only function that language can serve. This observation has been made by a wide range of theorists (for example, Chomsky, 1979) and can be confirmed by the layperson's introspective experiences. Whether it is in the conscious experience of thought as a silent speech stream or in the use of a rehearsal loop to bolster short term memory, we often recruit linguistic representations for non-communicative tasks. The perceived ubiquity of internal speech might suggest that the use of language in non-linguistic tasks is an inherent property of linguistic competence. If this is the case, we might expect to see evidence of widespread non-communicative language use from the earliest moments that an individual acquires fluent language skills. This paper explores this issue by contrasting 4-year-olds, 7-year-olds and adults on one test of non-communicative language use. Specifically, we looked at the propensity to spontaneously activate the phonological representation of object names in a visual memory task.

Young children do not report experiencing inner speech when asked about their mental processes during non-linguistic tasks (Flavell, Green, Flavell & Grossman, 1997; Flavell & Wong, 2009; Manfra, 2009). However, this could easily reflect a shortfall of introspection, rather than the absence of internal verbal representations. Thus, it is necessary to explore other approaches to the study of implicit language use. Since working memory in adults relies on phonological representations, even when the stimuli are visual,

---

<sup>1</sup> The studies described in this dissertation were all conducted in collaboration with Jesse Snedeker.

(Baddeley & Hitch, 1974), this domain is an attractive one for investigating the development of inner speech as a cognitive tool.

Studies on the development of working memory for visual stimuli have repeatedly failed to find evidence for verbal memory strategies in preschool aged children. Phonological similarity between items in a memory set does not impede working memory until children are school aged (Conrad, 1971; Hulme, 1984; Ford, & Silber, 1994), but visual similarity has an effect from an early age (Brown, 1977; Hayes & Schulze, 1977). In a paired-associate memory paradigm, Cramer (1975) found evidence of implicit labeling of pictorial stimuli in fourth and sixth grade children but not second graders. For example, the older children were more likely to show false recognition of a *penny* when *pen* and *knee* were paired in the memory set. These findings suggest that the younger children were not responding based on a verbal code. Similarly, sensitivity to word length and articulatory suppression in short term memory for pictures does not appear to emerge until seven or eight years of age, again indicating that younger children do not rely on verbal memory representations for visual stimuli (Allik & Seigal, 1976; Ford & Silber, 1994; Hitch & Halliday, 1983; Hitch, Halliday, Dodd, & Littler, 1989; but see Henry, Turner, Smith, & Leather, 2000; Hulme, Silvester, Smith, & Muir, 1986).

However, these results are not conclusive with respect to verbal encoding. Memory tasks with a delay present two opportunities for phonological effects to arise: 1) at encoding, participants may or may not retrieve labels for images; 2) during the delay, participants may or may not engage in verbal rehearsal (Locke & Fehr, 1970; Morrison & Haith, 1976). There is ample evidence that preschool aged children do not reliably use verbal rehearsal (Flavell, Beach & Chinsky, 1966; Gathercole, Adams, & Hitch, 1994;

Ornstein, Naus, & Liberty, 1975; but see Henry, 1991). When young children are explicitly instructed to rehearse memory items, both phonological similarity and word length effects do emerge, suggesting that their absence in other studies might be due to a lack of rehearsal (Johnston & Conning, 1990; Johnston, Johnson, & Gray, 1987). Thus, it is possible that children of all ages activate verbal codes for the to-be-remembered material, but only the older children strategically employ rehearsal, and it is the lack of rehearsal that accounts for the absence of effects of phonological similarity, word length and articulatory suppression in younger age groups. Some support for this perspective comes from the finding that 5-year-olds make silent lip movements, perhaps reflecting verbal encoding, during stimulus presentation but not during the retention interval (Locke & Fehr, 1970; but see Flavell, Beach & Chinsky, 1966). On the other hand, the observation that verbal rehearsal generally emerges later for pictorial stimuli than verbal stimuli (e.g. Hitch, Halliday, Dodd, & Littler, 1989; Hitch, Woodin, & Baker, 1989; Hulme & Tordoff, 1989; Johnston, Johnson, & Gray, 1987; but see Henry, Turner, Smith, & Leather, 2000) does suggest that young children are less likely to verbally encode visual material. However, this evidence is indirect and open to other explanations, such as verbal material triggering a rehearsal strategy (Macleod, & Posner, 1984). To address these issues, it is important to observe encoding directly, rather than drawing inferences from recall or recognition performance.

Several different lines of work have examined verbal encoding during visual tasks in adults. Meyer and colleagues (2007) found evidence for object label retrieval during a visual search task by measuring eye movement patterns as participants looked for the target image. Distractors that depicted objects that were homophonous with the label of

the target image attracted more looks than unrelated distractors, even though the task itself did not involve the object labels. Further evidence for verbal encoding comes from a study by Noizet and Pynte (1976) in which participants were instructed to silently look at images until they identified the depicted objects. Participants spent longer looking at objects with longer (multisyllabic) labels than those with shorter (monosyllabic) labels, indicating that they might be retrieving the verbal labels, even though they completed the study in silence. This finding, however, is open to alternate explanations: name length could potentially be confounded with visual or conceptual complexity, or participants might have interpreted the instruction to identify the images as an instruction to silently label them. Zelinsky and Murphy (2000) addressed these issues in a short-term visual memory task that also used a name length effect as a window into verbal encoding. In this study, participants were shown displays of four images to remember, followed by a 2.5 second pause, and then a probe image that they identified as old or new. The four to-be-remembered images consisted of two monosyllabic items and two multisyllabic items. They controlled for differences in conceptual and visual complexity by equating the stimuli based on categorization speed in a norming task. Zelinsky and Murphy found that adults spent longer looking at the multisyllabic items, indicating that they were activating verbal labels for the images during encoding. A second experiment with novel image-label pairs confirmed this finding.

The current study adapts the short-term visual memory paradigm from Zelinsky and Murphy (2000) to look at the development of verbal encoding in a visual memory task. Critically, unlike many of the previous developmental studies, our approach looks at verbal

label activation during encoding, rather than testing for verbalization effects at retrieval, at which point a failure to rehearse could mask the effects of verbal encoding.

We tested three age groups of participants in the current study: 4-year-olds, 7-year-olds and adults. These age groups were chosen for the following reasons. Four-year-old children typically do not show phonological rehearsal effects in the short-term memory literature. Seven-year-old children are on the cusp of showing such effects (e.g. Bach & Underwood, 1970; Ford & Silber, 1994; but see Cramer, 1975) and generally demonstrate more use of metacognitive strategies, perhaps due to experience with formal education (Miller, 1993). Finally, testing adults allows us to replicate the name-length effects from Zelinsky and Murphy (2000) and provides a baseline for evaluating the results from the younger age groups. In Experiment 1, we validate the paradigm as a test of verbal encoding in each of the age groups. Experiment 2 then uses this paradigm to investigate spontaneous verbal encoding across these age groups.

## **Experiment 1**

These studies employ a recognition memory paradigm based on Zelinsky and Murphy (2000). Before using this task with 4- and 7-year-olds, we needed to make sure that eye movements during the encoding phase are sensitive to verbal labeling in these age groups. In Experiment 1, participants named the images shown on the screen out loud, while encoding them for the memory task. Looking longer to images with longer labels in Experiment 1 confirms the sensitivity of the paradigm to labeling and allows us to interpret the presence or absence of implicit labeling in Experiment 2, which omits the labeling instruction.

## **Method**

### *Participants*

Our participants consisted of 15 4-year-old children (4;3 to 4;11 years,  $M=4;8$  years, ten females), 15 7-year-olds (7;1 to 8;0 years,  $M=7;6$  years, five females), and 15 adults (18 to 50 years,  $M=28$ , seven females). Children were recruited from a database of families who had expressed interest in participating in research. The adults were drawn from the Harvard University participant pool, which includes students and members of the community. Five additional 4-year-olds participated in this study but were excluded from these analyses, one due to poor accuracy on the memory task, and four due to extensive eye tracking data loss. For both Experiment 1 and Experiment 2, all participants were native English speakers. Also, in both experiments, children were given a small toy as a prize for participation and adults were compensated with partial course credit or a payment of \$5.

### *Materials*

Twenty colored line drawings were selected: 19 from the Rossion and Pourtois (2004) image set and one stylistically similar picture from an internet image search. Half the images had monosyllabic labels and the other half had three- or four-syllable labels compiled (see Appendix A for a complete item list). Since name length can correlate with visual complexity (Kelly, Springer, & Keil, 1990), we conducted a visual-categorization speed norming task.. In this procedure, taken from Murphy and Brownell (1985) and Zelinsky and Murphy (2000), participants saw a printed word on-screen immediately followed by an image, at which point they made a speeded judgment about whether the picture matched the word. Since name retrieval is not necessary at the time the participant sees the image, this task allows us to isolate differences in reaction time based on visual

complexity. For the pictures used in these experiments, the average reaction times for the monosyllabic and multisyllabic items were 608ms and 553ms respectively (by subjects:  $t(9) = 2.27, p < 0.05$ , by items:  $t(18) = 2.40, p < 0.05$ ). It is important to note that the difference in reaction times is in the opposite direction to the predicted name length effect, so visual recognition processes should not provide a competing explanation for greater looking time to these items. We also collected visual complexity ratings for the study images from participants on Amazon Mechanical Turk. Each image was rated on a scale from 1 to 7 (*very simple* to *quite complex*) by 10 native English speakers. The ratings for multisyllabic and monosyllabic images were not significantly different (Mann-Whitney  $U=58, p=.579$ ).

We determined the frequency of the object labels in the CHILDES database (MacWhinney, 2000), using the ChildFreq tool (Bååth, 2010). In transcripts in which the target child was 4-years old, the monosyllabic words occurred, on average, 110 times per million words (range: 14 to 292 per million words) and the multisyllabic words occurred 70 (range: 5 to 216) times per million words ( $t(18)=1.19, p>.2$ ). In transcripts from 7-year-olds these values were 80 (range: 0 to 366) and 169 (range: 0 to 889) occurrences per million, respectively ( $t(18)=0.85, p>.4$ ).

#### *Procedure and apparatus*

The experiment was conducted using a Tobii T60 screen-based eye tracker, with a sampling rate of 60Hz. Each session began with a calibration for the eye tracker using the Clearview software. After an acceptable calibration was obtained, the experiment was run in E-Prime.

Each trial began with a blue circle in the center of the screen. After the participant fixated on the circle for 500ms, it disappeared and the encoding phase of the trial began. The to-be-remembered display consisted of four images, one in each quadrant of the screen. This display stayed on-screen for five seconds. Participants were instructed that during this part of the trial they should label each of the images out loud. If children failed to do this on the practice trials, the experimenter prompted them to name the images while they were still on the screen. During the experimental trials, all adults and most children labeled the objects without prompting or feedback. If a child did not name the objects out loud, then the experimenter reminded her to label the images before the next trial. The five-second encoding phase was immediately followed by the test image. The participants' task was to indicate whether the test image had been present in the four picture display which they had been asked to remember. Adult participants responded with a key press, while children gave verbal responses ("yes" or "no"). Responses were followed by visual feedback, a smiley face or a cross, which was accompanied by auditory feedback for the children ("nice job" or "better luck next time"). As incentives, children were given stickers for every ten correct responses. There were four practice trials and forty experimental trials.

In each to-be-remembered picture display, there were two monosyllabic items and two multisyllabic items. Each image appeared eight times over the course of the experiment, twice in each location, and no pair of images appeared in the same display together more than twice. The test image was always taken from the set of 20 images that were used across trials in the encoding phase, so participants could not respond based on familiarity alone. On half of the trials the test image had been present in the to-be-



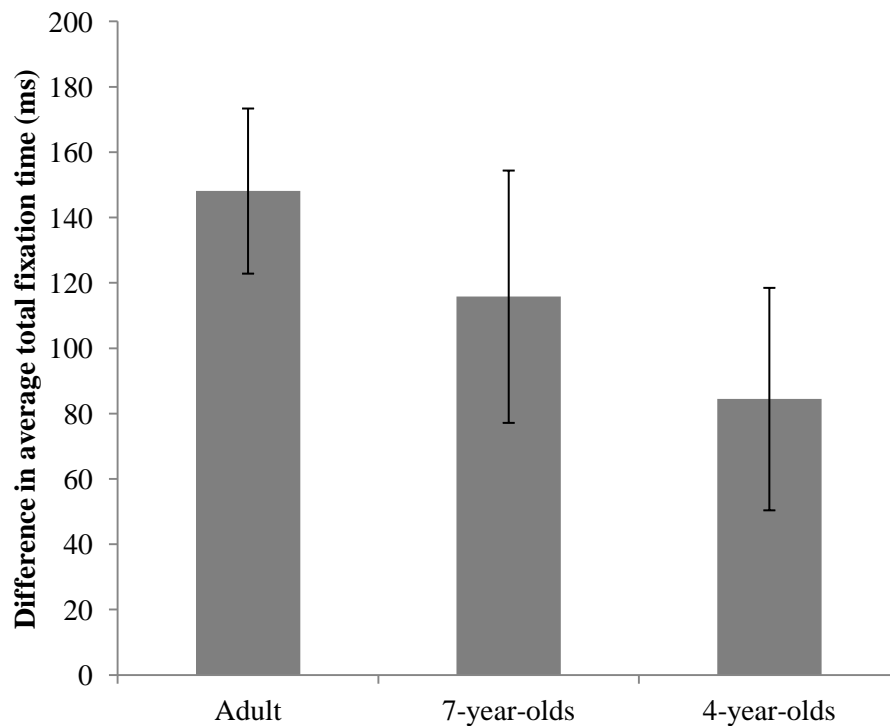
remembered display and on half it had not. Trial order was randomized for each participant. After completing the memory task, the children were shown each image again and asked to label it, to determine whether they associated the images with the expected names. We did not conduct the naming post-test with adults since 19 of the 20 images came from a database of images that had already been normed for name agreement in adults (Rossion & Pourtois, 2004)

## **Results and Discussion**

An accuracy criterion of 80% correct responses was used for inclusion in this study; one 4-year-old did not meet this standard. Including all children, 4-year-olds' average percentage of correct responses was 88.4%. After removing the score for the excluded participant, this percentage increased to 90.1%. Seven-year-olds' average accuracy was 97.6% and adults' average accuracy was 97.7%. An analysis of variance on the log-odds transformed accuracy scores revealed a significant effect of age on accuracy,  $F(2,42) = 13.41, p < 0.001$ . In light of this effect, we only analyzed eye gaze for trials with correct responses.

We calculated total fixation time to each image during the five-second encoding phase by summing the fixations recorded for each quadrant. The validity metric provided by E-prime Extensions for Tobii was used to filter out samples that did not have valid measurements from both eyes. We also removed samples that recorded a gaze location in a different quadrant than both the sample before and the sample after, since such short fixation durations were likely to reflect measurement errors. Finally, if during the five second display, the total fixation time for an on-screen image was less than 150ms, the fixation time to that image was not considered in our analyses. Based on performance

profiles in various rapid picture identification paradigms (Fize, Fabre-Thorpe, Richard, Doyon, & Thorpe, 2005, Kirchner, & Thorpe, 2006; Potter, 1976; Potter, & Levy, 1969), we reasoned that if an image was fixated for less than this amount, the eye tracking data likely did not reflect substantial object processing, and that the image was either encoded extrafoveally or was not encoded at all.



*Figure 1.* Experiment 1 total fixation times. Difference between the average total fixation time to the multisyllabic and monosyllabic items in each display when participants were explicitly labeling each image. Error bars display the standard error of the difference score.

For each trial, we calculated the difference in the average total fixation time to the monosyllabic items and the multisyllabic items, excluding items that were fixated for less than 150ms (Figure 1). We analyzed the fixation time difference score using a series of mixed effects regressions, implemented using the lme4 library (Bates, & Maechler, 2010) in R, version 2.11.1 (R Development Core Team, 2010); *p* values were estimated assuming the

$t$  values were drawn from a normal distribution, using the `pnorm` function. First, we conducted a mixed-effects linear regression with age group as a fixed effect, including random intercepts for subjects, and using Adults as the reference age group (Table 1). In this model, we found that the intercept was significantly greater than 0 ( $p < .01$ ), indicating that adults looked longer to the multisyllabic items, and there were no significant effects of age ( $ps > .2$ ). We conducted separate regressions for each age group to confirm that the name length effect occurred for the younger age groups as well. At each age level, the intercept was significantly greater than 0 ( $ps < .05$ ), confirming participants looked longer to images with multisyllabic labels when naming them explicitly.

*Table 1.* Fixed effects from mixed-effects linear regression models of total fixation time difference (multisyllabic – monosyllabic) with subjects as random effects. Intercepts significantly greater than 0 indicate a name length effects for that age group (\* $p < .05$ , \*\*\* $p < .001$ ).

Age Group	Fixed Effect	Estimate	SE	$t$ value
All (Adults as reference)	(Intercept)	151.83	30.70	4.945***
	Seven year olds	-44.31	44.33	-1.000
	Four year olds	-69.53	45.69	-1.522
Adults	(Intercept)	151.83	28.77	5.277***
Seven year olds	(Intercept)	107.52	29.88	3.599***
Four year olds	(Intercept)	82.30	38.44	2.141*

The findings from this control experiment confirm that the eye movement patterns are sensitive to naming in this task, setting the stage for investigating implicit naming during image encoding.

## **Experiment 2**

In Experiment 2, we tested whether the name length effect would still be observed in the absence of an explicit naming instruction. For the adult subjects, we aimed to replicate Zelinsky and Murphy's (2000) finding that adults look longer at pictures with longer names. By extending this design to 4- and 7-year-olds, we can determine the developmental trajectory of implicit labeling in this task.

## **Method**

### *Participants*

As before, three age groups of participants were included in this experiment. Twenty 4-year-olds (4;5 to 4;10 years,  $M=4;9$  years, 11 females), 20 7-year-olds (7;1 to 8;0 years,  $M=7;8$  years, nine females), and 20 adults (18 to 59 years,  $M=32$  years, 13 females) participated. Six additional 4-year-olds participated in this study but were excluded from these analyses: five due to poor accuracy on the memory task and one for not following the task instructions. Also, data from two additional adult participants could not be used due to poor eye tracking.

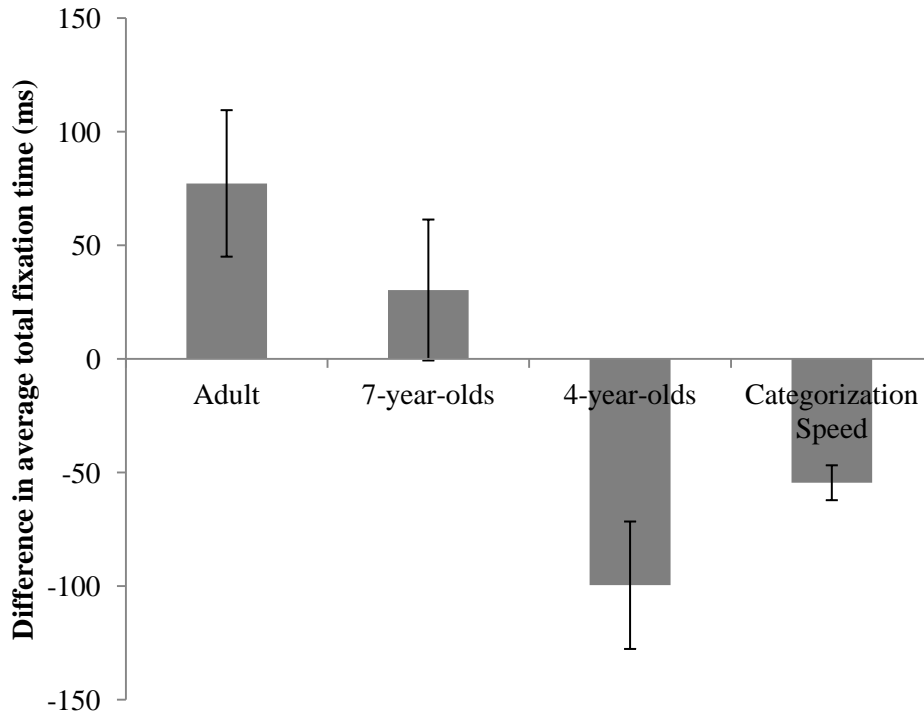
### *Materials, procedure, and apparatus*

The methods were identical to Experiment 1, except that participants were not told to name pictures out loud during the encoding phase of the trial, that is, the five second period while the to-be-remembered display was on the screen.

## Results and Discussion

As in Experiment 1, we used an 80% accuracy criterion for inclusion in the study; five 4-year-olds did not meet this standard. Including all participants, 4-year-olds' average correct response rate was 80.9%, after removing the scores for the five participants who did not meet the criterion, the average correct response rate was 85.4%. The average correct response rate for adults was 94.0% and for 7-year-olds was 93.2%. An analysis of variance on log-odds transformed accuracy scores indicated that there was a significant effect of age,  $F(2,57)=14.83$ ,  $p<0.001$ . Accuracy in Experiment 2 was substantially lower than in Experiment 1 ( $F(1,99) = 15.56$ ,  $p < 0.001$ ), and this effect did not interact with age,  $F(2,99) = 0.10$ ,  $p>.9$ . As in Experiment 1, only trials with correct responses were included in the eye-gaze analysis.

As we did in Experiment 1, we calculated the difference between average total fixation time to the multisyllabic and monosyllabic images on each trial, excluding images with total fixation time of less than 150ms (Figure 2). We first conducted a mixed-effects linear regression with age group as a fixed effect and including random intercepts for subjects, and using Adults as the reference age group (Table 2). The analysis revealed a significant name length effect for the adults, as shown by the intercept ( $t=2.48$ ,  $p<.05$ ), and a significant difference in the name length effect between adults and 4-year-olds ( $t=-3.87$ ,  $p<.001$ ).



*Figure 2.* Experiment 2 total fixation times. Difference between the average total fixation time to the multisyllabic and monosyllabic items in each display when participants were not instructed to label the images. The final bar depicts the average difference in categorization speed (from the norming task described in Experiment 1) for these images. Error bars display the standard error of the difference score.

We also conducted separate regressions for adults, 7-year-olds and 4-year-olds to investigate the possibility of implicit naming at each age level. Again, intercepts significantly greater than 0 indicated that participants looked longer at the images with multisyllabic labels, and so were likely to be retrieving object labels during the task. The expected name length effect was confirmed for the adults (Intercept=76.78ms,  $t=2.51$ ,  $p<0.05$ ), but was not observed in the younger age groups. In fact, the 4-year-olds looked significantly longer at the images of objects with *shorter labels* (Intercept=-101.11ms,  $t=2.77$ ,  $p<0.01$ ).

*Table 2.* Fixed effects from a mixed-effects linear regression model of total fixation time difference (multisyllabic – monosyllabic) with subjects as random effects. Intercepts significantly greater than 0 indicate a name length effects for that age group (\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ ).

Age Group	Fixed Effect	Estimate	SE	<i>t</i> value
All (Adults as reference)				
	(Intercept)	76.78	30.88	2.486*
	Seven year olds	-49.95	44.21	-1.130
	Four year olds	-177.89	46.02	-3.865***
Adults				
	(Intercept)	76.78	30.57	2.512*
Seven year olds				
	(Intercept)	26.83	29.97	0.895
Four year olds				
	(Intercept)	-101.11	36.51	-2.770**

To explore this unexpected reversal of the name length effect, we turned to the categorization speed norming task that we had conducted (described under Materials in Experiment 1). In this norming task, we found that participants actually responded more quickly to the monosyllabic visual stimuli than the multisyllabic items. If these differences in visual categorization had a larger effect on children’s fixation times, perhaps because of a reliance on visual memory strategies, then this would explain the reverse name length effect in 4-year-olds. Thus, we calculated the difference in average categorization time between the monosyllabic and multisyllabic items in each display and included this categorization speed difference score as a predictor for the fixation time difference score. We conducted a mixed-effects regression including age group and categorization speed difference score as fixed effects, with random intercepts for subjects and display combinations, including random slopes where appropriate (Table 3). This analysis yielded

an intercept of 161.67ms ( $t=2.43, p<.05$ ), again indicating that adults looked longer at images of objects with longer names (Table 5). The categorization speed difference score was a marginally significant predictor overall ( $t=1.83, p=.07$ ). Further, there were significant interactions of this predictor with age, for both 4- and 7-year-olds ( $ts>2, ps<.05$ ), indicating that this index of visual complexity predicted fixation times differently for the different age groups.

As before, we conducted separate regressions for each age group, now including the categorization speed difference score as a predictor (Table 3). In adults, the naming effect was observed ( $t=2.00, p<.05$ ) and there was no significant effect of visual complexity, as measured by categorization speed. Seven-year-olds showed both a naming effect (Intercept=187.75ms,  $t=2.64, p<0.01$ ), and a visual complexity effect ( $t=3.09, p<.01$ ), indicating that they were likely to be implicitly labeling the images. Once visual complexity was accounted for, fixation times in 4-year-olds no longer showed any name length effects ( $t=0.87, p>.3$ ), but did show an effect of visual complexity ( $t=2.93, p<.01$ ). These analyses were confirmed by conducting model comparisons, using ANOVAs, to assess whether models that include the visual complexity predictor explained significantly more variance than those without this predictor. For the four and 7-year-olds, including the visual complexity predictor explained significantly more of the variance ( $ps<.05$ ), but this was not the case for the adults ( $p>.4$ ).

These data provide evidence of spontaneous name retrieval during the encoding of pictures into short term memory for adults and 7-year-olds, but not 4-year-olds.



*Table 3.* Fixed effects from a mixed-effects linear regression model of total fixation time difference (multisyllabic – monosyllabic) including categorization speed difference score as a predictor and with subjects and display combinations as random effects. Intercepts significantly greater than 0 indicate a name length effect for that age group, after accounting for differences in categorization speed across items ( $^{\wedge}p<.07$ ,  $*p<.05$ ,  $**p<.01$ ).

Age Group	Fixed Effect	Estimate	SE	t value
All (Adults as reference)	(Intercept)	161.67	66.39	2.435*
	Categorization Speed Difference Score	1.57	0.86	1.826 $^{\wedge}$
	Seven year olds	64.94	67.74	0.959
	Four year olds	-52.86	77.73	-0.680
	Categorization Speed Diff. Score x Seven y.o.	2.14	0.92	2.337*
	Categorization Speed Diff. Score x Four y.o.	2.17	1.02	2.116*
Adults	(Intercept)	141.83	70.87	2.001*
	Categorization Speed Diff. Score	1.20	0.96	1.247
Seven year olds	(Intercept)	187.75	71.26	2.635* *
	Categorization Speed Diff. Score	2.98	0.96	3.092* *
Four year olds	(Intercept)	69.76	80.68	0.865
	Categorization Speed Diff. Score	3.01	1.03	2.930* *

## General Discussion

The results of these eye tracking experiments demonstrate that there is a developmental change in the use of verbal encoding for working memory tasks during the

early school years. We found that 7-year-olds, like adults, spontaneously engaged in verbal encoding in a short-term visual memory task. In contrast, 4-year-olds did not verbally encode the pictures, unless they were specifically instructed to do so. Experiment 1 established that our fixation time index was in fact sensitive to naming in all three age groups, strengthening the conclusion that 4-year-olds were not activating verbal labels in Experiment 2. The age differences in name length effects, in combination with the visual complexity effects, suggest a developmental trajectory in which children begin with visual encoding strategies for short-term picture memory. Children start combining this with verbal encoding after entering school, and adults show an increased reliance on verbal labels. Below we discuss these findings in the context of prior studies on working memory development and implicit verbal encoding.

The results of this study are consistent with earlier work on the development of working memory, which has often shown that verbal memory strategies are not used for visual stimuli until the elementary school years. However, many of these prior findings relate to verbal rehearsal rather than verbal encoding, for instance, the repeated finding that verbal rehearsal emerges earlier for verbal rather than visual stimuli (e.g. Hitch, et al., 1989; Hulme & Tordoff, 1989; Johnston, et al., 1987; but see Henry, et al., 2000). It is important to note that verbal rehearsal for visual stimuli necessarily depends on the emergence of verbal encoding. The current findings suggest that this tendency to recode visual stimuli into verbal representations is not present in preschoolers, offering one explanation for the absence of verbal rehearsal.

Of course, it is also possible that verbal encoding in the absence of verbal rehearsal is not a particularly useful cognitive approach, and so it might be the acquisition of

rehearsal strategies that is the limiting factor for the emergence of verbal encoding. However, Macleod and colleagues found that the act of overtly verbally encoding the stimuli led to improved memory performance, compared to silent encoding (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). They suggest that labeling aids memory because it increases the distinctiveness of the representation. Similarly, in our study, memory performance improved for all age groups when they overtly labeled the images in the display, again suggesting that there might be benefits just from verbal encoding. The verbal encoding advantage becomes a disadvantage if there are multiple exemplars corresponding to the same label, e.g. a study set including several images of chairs (Lupyan, 2008), again indicating that verbal processes at encoding can affect memory performance independently of rehearsal.

Our results are also consistent with prior developmental work on the use of visual and linguistic representations in conceptual tasks. In a developmental study of visual imagery, Kosslyn (1976) found that adults were slower to perform a conceptual judgment when they were instructed to perform the task using visual imagery, compared to when that instruction was omitted. This suggests that visual imagery was not the default mode of representation for the adults in this study. On the other hand, first graders performed similarly whether or not they were instructed to use visual imagery. These results suggest that young children's default encoding strategy might be more image-based than adults', which is consistent with the current study.

On the other hand, findings from a different paradigm suggest that even younger children, 18- and 24-month-old infants, do engage in verbal encoding for visual images. In a series of priming studies, Mani and colleagues found that viewing an image in silence can

effect a infants subsequent ability to process a word that is phonologically related to that image, even though the label of the original image was not spoken or heard (Mani & Plunkett, 2010; 2011; Mani, Durrant, & Floccia, 2012). Building on these results, we found that 24-month-old infants spontaneously activated the phonological form of image labels even in a completely non-communicative context without any explicit verbal labeling (Khan, Geojo, Wang, & Snedeker, submitted).

How can we account for the divergent pattern of development across tasks? Perhaps the diverse results across age groups can be reconciled by viewing verbal encoding in terms of its function in the task at hand. The studies that show implicit naming in 18- and 24-month-olds require infants to attend to only one or two images at a time, and these images depict objects with which they are highly familiar (Khan, et al, submitted; Mani & Plunkett, 2010; 2011; Mani, et al., 2012). This context is very similar to the naming games that infants this age play with caregivers (“What’s that?”). The critical difference, of course, is that in these studies the game is silent, spontaneous and self-directed. Infants at this age are actively engaged in word learning, often through ostensive labeling of visual objects. Thus, perhaps it is functional for them to engage in verbal processing in these kinds of contexts. Naming games of this sort might provide an opportunity for infants to rehearse and consolidate word-meaning pairs.

In contrast, the current study uses a recognition memory task as a vehicle for investigating inner speech in 4-year-olds. While memory tasks appear to strongly elicit verbal encoding in adults, this may be a context in which verbal encoding is late developing. Verbal encoding as a memory strategy requires a sophisticated understanding of the goals of memory tasks and of how memory works. Even though verbal encoding aids memory

performance in young children, they might not spontaneously use a labeling strategy unless they are explicitly aware of its benefits. Monitoring the efficacy of a memory strategy appears to be late developing (Pressley, Levin, & Ghatala, 1984), and preschoolers do not have the benefit of formal instruction of memory strategies, so it is likely that 4-year-olds do not realize memory performance benefits from verbal encoding. Prior to this realization, there would be little incentive for a young child to devote resources to verbally recoding visual material in a memory task.

While there are many issues to be resolved, the current findings provide a strong and novel constraint on our understanding of how working memory develops in early to middle childhood. Specifically, it demonstrates that preschoolers not only fail to verbally rehearse but in fact do not even verbally encode visual stimuli that they are trying to remember. In contrast, young school-aged children do implicitly verbally encode images in a working memory task, as do adults. This suggests a developmental change in the mechanisms supporting working memory in the early school years, perhaps due to increases in metamemory following formal education (see Pressley, Borkowski, & Sullivan, 1985; Schneider & Lockl, 2002; Schneider & Pressley, 1997, for reviews). The presence or absence of implicit verbal representations during encoding can be detected using eye-tracking. This tool can be used to assess the role of linguistic representations across a range of cognitive tasks and age groups to gain a more complete picture of the development of verbal encoding strategies.

## Chapter 3. Spontaneous verbal labeling in 24-month-old infants<sup>2</sup>

A central and controversial issue in cognitive science is the role of language in internal thought. Theorists as diverse as Mead (1934), Vygotsky (1934/1987), Whorf (1956) and Chomsky (2007) have proposed that language is the primary vehicle for internal thought. In contrast, others have argued that thought must occur over representations that are distinct from natural languages (Fodor, 1975; Jackendoff, 2002; Pinker, 1994). In some cases, the diverging opinions reflect differences in how the word “language” is defined—whether it picks out conceptual and semantic representations, syntax, phonology, or all of the above (see Jackendoff, 2002). Nevertheless, after definitional differences are sorted out and evolutionary claims are put aside, a clear empirical question remains: Does the specific form of the external language that we use to communicate play a role in our internal cognitive life? This question is often approached by finding languages with diverging syntactic and semantic categories and looking for differences in the performance of non-communicative tasks (see Gentner & Goldin-Meadow, 2003 for examples), but it can also be addressed by testing whether the phonological forms of linguistic expressions are active during nonlinguistic tasks.

There is no doubt that internal verbalization occurs. We have mental conversations with people who are absent, we silently coach ourselves during difficult tasks (“Eyes on the ball!”), and we use subvocal rehearsal as a short-term memory aid (Martínez Manrique & Vicente, 2010; Vicente & Martínez Manrique, 2011; Winsler, 2009). But several questions remain. How ubiquitous is verbal encoding: is it largely limited to strategic, metacognitive contexts like those above, or does it pervade our spontaneous thought? Does verbal

---

<sup>2</sup> The study in this chapter was conducted in collaboration with Amy Geojo, Shanshan Wang and Jesse Snedeker.

encoding typically involve the activation of phonological representations or does internal thought use the semantic representations of language stripped free from their external trappings? And finally, when in development does spontaneous verbal encoding, outside of communicative contexts, first emerge? The present paper begins to explore this last question, by looking at the activation of verbal labels for visually-presented objects in 24-month-old infants.

While there is some evidence that adults spontaneously activate phonological representations in noncommunicative contexts, the data pattern is complex and ambiguous. In recognition memory tasks, the time spent looking at an object during the encoding phase is affected by name length of the object (Zelinsky & Murphy, 2000). Although this task had only a 2.5 second delay and could be solved visually, this pattern indicates that adults encoded the pictures verbally, perhaps so they could verbally rehearse their labels. Critically, the authors found no effect of name length in a parallel visual-search task which had no memory demands. There is, however, some evidence for the activation of linguistic form during visual search. Specifically, the presence of homophonous competitors results in interference (e.g., more looks to a baseball bat when searching for an animal bat) suggesting that the lexical label of the picture has been retrieved (Meyer, Belke, Telling, & Humphreys, 2007). Curiously, visual search tasks are not influenced by phonological overlap between the target and distractors; persons searching for “candy” do not linger on candles (Telling, 2008). This raises the possibility that the homophone effects either reflect activation at the lexical level (rather than at the phonological level) or depend on the participants’ metalinguistic awareness of homophony.

The origin of internal verbal representations has been a central theoretical issue in developmental psychology. Vygotsky (1934/1987) proposed that children initially use language solely as a social tool, with internal speech developing gradually from external dialog. The first step toward internalization occurs at around two or three when children begin talking aloud to themselves. This private speech becomes quieter, less frequent, more elliptical, and more covert during the late preschool and early elementary years (Berk, 1986; Winsler & Naglieri, 2003; see Winsler, 2009 for review). Researchers working within the Vygotskyian tradition interpret these changes in private speech as evidence for the emergence of internalized verbal thought (see e.g., Fernyhough, 2009).

However, there is little direct evidence that internal verbalization is absent before the rise and fall of private speech. Unlike adults, young children do not report thinking in words (Flavell, Green, Flavell, & Grossman, 1997; Manfra & Winsler, 2006), but this could reflect a failure to understand the question or introspect, rather than the absence of internal verbalization (Flavell & Wong, 2009). In memory studies with pictorial stimuli, preschoolers make errors based on visual similarity but do not show effects of word length, phonological similarity or verbal suppression, suggesting that they do not use internal verbalization even in those tasks where adults rely on it (Conrad, 1971; Ford & Silber, 1994; Hayes & Schulze, 1977; Hitch & Halliday, 1983; Hitch, Halliday, Dodd, & Littler, 1989; but see Henry, Tuner, Smith, & Leather, 2000; Hulme, Silvester, Smith, & Muir, 1986). This pattern, however, could reflect a failure to use verbal rehearsal or retrieval strategies, rather than the absence of verbal encoding (Hitch & Halliday, 1983; Johnston & Conning, 1990; Nairne, 1990). In fact, in picture memory studies 5-year-olds have been observed making silent lip movements as the pictures are presented, suggesting that they



are retrieving their verbal labels (Locke & Fehr, 1970; but see Flavell, Beach, & Chinsky, 1966).

In this paper, we explore a fundamentally different hypothesis about the development of internal verbalization, one that is rooted in psycholinguistic research and the information-processing tradition. Psycholinguists construe language as a series of linked representations (phonological, lexical, syntactic and semantic) which are constructed during comprehension and production (see e.g., Alario, Costa, Ferreira, & Pickering, 2006; Altmann, 2001; Snedeker, 2009). Language development involves, among other things, acquiring connections between levels, such as the mappings between phonological word forms and the concept that a word expresses (Jackendoff, 2002). Once a connection has been acquired, activation at one level of representation can result in activation of the linked representation, as it must for successful word comprehension or production. Covert speech occurs when lexical and phonological representations are constructed without giving rise to articulatory plans (Indefrey & Levelt, 2004). On this construal, the representational basis for internal verbalization is available as soon as language acquisition begins, raising the possibility that older infants might already think in words. To explore this prediction, we looked for evidence of phonological activation in 24-month-olds who were passively viewing pictures.

The present study builds directly on experiments by Mani and Plunkett (2010; 2011; Mani, 2010) which show that infant word recognition is sensitive to phonological and phono-semantic priming. For example, in the phonological priming study, infants were shown a prime image (e.g. a picture of a cup), unaccompanied by an explicit label, followed by split screen with two images (a cat and a house), one of which was explicitly labeled

("cat"). Infants' looking times to the target image (cat) were affected by whether or not the label is phonologically related to the name of the previous silent image: 18-month-olds show phonological facilitation, while 24-month-olds show phonological interference.

For these effects to emerge, infants had to activate the label of the prime image, which was never spoken aloud. Thus some form of internal verbal activity occurred. Mani and Plunkett (2010) interpret this activity as implicit naming, and draw a parallel to the homophone effects observed in adults by Meyer and colleagues (2007). However, this task is different from Meyer's in a fundamental way: the unlabeled prime is embedded in an overt word-recognition task, rather than a non-linguistic search task. Throughout the experiment infants heard an alternating stream of labeled targets and unlabeled primes. Consequently, infants may have generated labels for primes because they expected to hear these words or because they understood this task as a labeling game and were trying to play along. In fact, in adults, unlabeled pictures produce phonological interference when they are embedded in a linguistic task, like picture naming (Meyer & Damian, 2007; Morsella & Miozzo, 2002), but not when they are embedded in a nonlinguistic task like visual search (Telling, 2008; see also Zelinsky & Murphy, 2000). Thus, the prior preferential-looking studies do not answer the core question that motivates the present experiment: Do infants verbalize their experiences in non-linguistic contexts, or do they do so only in expectation of hearing or producing external labels?

The current study addresses this question by looking for phono-semantic priming in 24-month-olds in a non-linguistic task in which objects are never labeled for the child and the child is never asked to produce a label. In phono-semantic priming, the prime is phonologically related to an intermediary that is semantically related to the target; the

intermediary is never explicitly presented during the task (e.g. *cup-cat-dog*). The phenomenon of phono-semantic priming has been observed both in adult and child language processing (Huang & Snedeker, 2011; Marslen-Wilson, 1987; Yee & Sedivy, 2006), and has been observed in 24-month-old infants using the preferential-looking priming paradigm described above (Mani, 2010; Mani, Durrant, & Floccia, 2012). Phono-semantic priming is an attractive tool for examining implicit phonological activation because phono-semantic relations are more difficult for participants to notice than phonological relationships, and therefore less available for strategic processes. Nevertheless, phono-semantic priming still relies on activating the phonological label of the prime.

The present study used a preferential-looking paradigm modeled on Mani (2010); infants saw a prime image followed by a split-screen of two images, one of which was phono-semantically related to the prime on half the trials. There was, however, one critical difference in our paradigm: none of the pictures (targets or primes) were labeled for the child. If infants' free-viewing of the split-screen differs based on whether or not one of the images is phono-semantically related to the previous image, this will provide strong evidence that phonological labels are spontaneously implicitly activated when infants encounter familiar visual objects. Such a finding would be consistent with the view that linguistic representations are recruited for internal thought, even in the early stages of language acquisition.

## **Method**

### *Participants*

Twenty four 24-month-olds from English speaking households participated in this study. An additional four infants were tested but could not be included in the final dataset:

one infant did not complete the experiment, and three were excluded due to technical failures or experimenter error.

### *Materials*

Ten prime-target-distractor triplets of color images were compiled (see Appendix B for a complete item list). In each triplet, the prime was phono-semantically related to the target; for example, *cup* is phono-semantically related to *dog* because the label “cup” shares its phonological onset with “cat”, and cats are semantically related to dogs. The intermediary or subprime, in this case *cat*, is never mentioned nor visually presented during the task. The distractor item was phonologically, semantically and phono-semantically unrelated to both the prime and the target. An unrelated control version of each item was created by shuffling the prime and target images between items, such that they were no longer phono-semantically related (and were also not phonologically or semantically related). Distractors were yoked to the target image during this shuffling. Since nothing is labeled during this task, an image is deemed the target simply by virtue of having been phono-semantically related to the prime in the original item triplets that were created. Each participant saw each item in either the related-prime or the unrelated-prime condition. Four experimental lists were created such that, for every item, related-prime and unrelated-prime conditions and target position on the screen were counterbalanced across participants.

The items used in this study were adapted from Mani (2010), with some substitutions due to differences between British and American English. The words that were relevant to the task (primes, targets, distractors and subprimes) were, on average, produced by 75% of 24-month-olds in the norming study for the MacArthur-Bates

Communicative Development Inventory (Dale & Fenson, 1996). Further, following the preferential-looking task, participants' knowledge of the words relevant to the task was confirmed by asking the caregiver to complete a vocabulary checklist, which indicated that these words were comprehended by an average of 97 percent of the infants tested (range across items = 88 percent – 100 percent). While caregivers completed this checklist, infants were asked to name the prime images that had been presented during the study. Critically, this naming task occurred after the preferential looking task. Images were presented in a booklet and infants were only prompted with questions such as “what is this?” or “can you help me name this?”. Five infants did not complete this task because they were too shy or tired. The pictures were correctly named by an average of 92% of the infants who completed this task (range = 84% – 100% by picture).

In order to rule out the possibility that direct semantic or visual similarity between the related-prime and the target could account for differences from the unrelated control condition, we collected adult ratings for these items on Amazon Mechanical Turk. For every item, the semantic relatedness between the related-prime and the target, the unrelated-prime and the target, and the related-subprime and the target were assessed with ten adults rating each pair on a scale from 1-7 (unrelated to highly related). The visual similarity between the related-prime and the target, and the unrelated-prime and the target, were also assessed. These ratings confirmed that neither the related and unrelated primes were semantically related to the targets; mean ratings were 1.54 and 1.33 respectively, and these ratings were not significantly different from each other,  $t(9)=0.96$ ,  $p>.3$ . Further, in both cases, the semantic relatedness ratings were significantly lower than for the subprime-target pairs ( $M=5.61$ ,  $ts>23$ ,  $ps<.001$ ). The visual similarity ratings did not

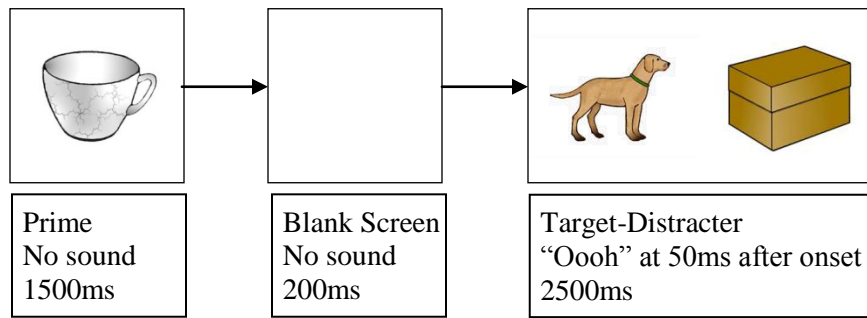
significantly differ between the related and unrelated prime conditions, mean ratings were 1.43 and 1.51 respectively,  $t(9)=0.34$ ,  $p>.7$ .

### *Procedure*

Participants sat on their caregiver's lap on a chair approximately 6 feet from the projector screen; caregivers were asked to keep their eyes closed for the duration of the task. A video camera immediately below the screen recorded the child's face so that their eye movements could be coded. Audio that was played through speakers behind the screen was also recorded to the videotape.

The experiment procedure was modeled on Mani (2010). Each participant viewed 5 trials in the related-prime condition and 5 trials in the unrelated-prime (control) condition with trial order randomized. The sequence of events within a trial is depicted in Figure 3. First, the participant saw the prime image presented in the center of the screen for 1500ms with no sound. Next, a blank screen appeared for 200ms, followed by the target and the distractor image. These images appeared on-screen side by side for 2550ms, with left-right order of the two pictures counterbalanced across items and experimental lists. 50ms after the target and distractor images appeared, a pre-recorded attention-getter in a female voice was played, such as "oh" or "ah". Finally, the participant saw a blank screen until the experimenter judged s/he was looking at the center of the screen, at which point the prime image for the next trial was presented. The related and unrelated prime conditions only differed with respect to the pairing of prime and target images.

Following the preferential-looking task, caregivers completed the vocabulary checklist. During this time, infants were shown the prime images again and asked to name them.



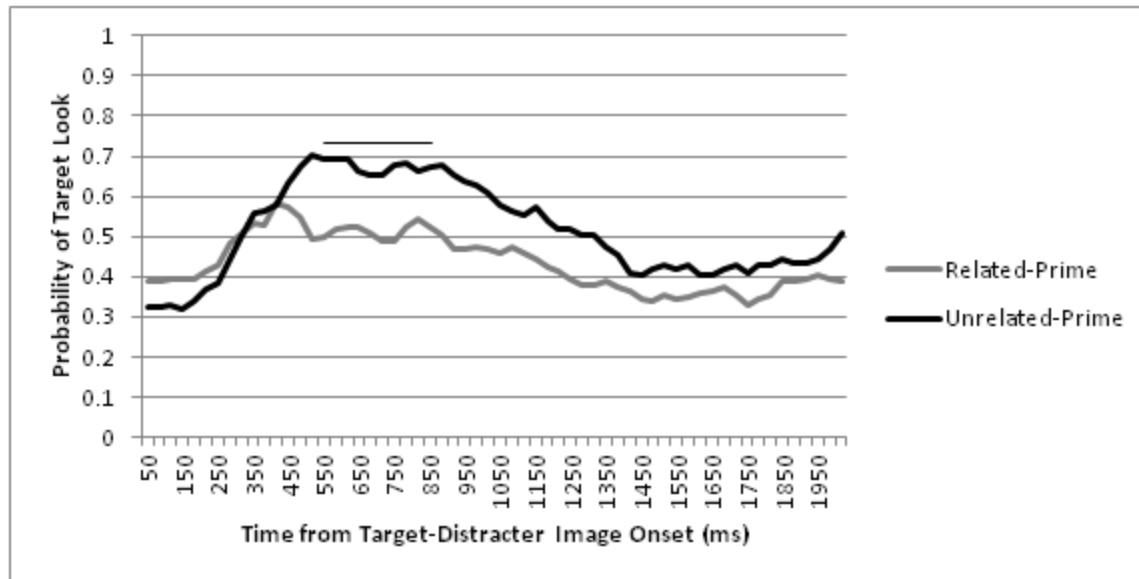
*Figure 3.* Schematic of a trial in related-prime condition (prime *cup*, subprime *cat*, and target item *dog*). In the unrelated-prime condition, the sequence remains the same except for the prime image, which would be replaced with a picture of a house.

### *Coding*

Eye movements were coded from the videotape using frame-by-frame viewing (33ms per frame). Fixations during the target-distracter screen were coded as left, right, center or away. Frames where the participant's pupils were not clearly visible were coded as track loss. Coding began 50ms after the target and distracter images appeared on-screen, at the point when the attention-getter was heard, and continued for 2000ms. Trials were excluded if the participant had not looked at the prime image or if there was more than 50% track loss during the critical time window (3.3% of trials). No participants were excluded for showing a side bias (> 80% looking to either the left or right side of the screen across the ten trials).

### **Results**

For each frame between 50ms and 2050ms from target-distracter image onset, we noted whether the participant was looking at the target for each trial (see Figure 4). Frames where the participant was not looking at either the target or the distracter were excluded from the analyses (5% of related-prime condition frames, 4% of unrelated-prime condition frames).



*Figure 4.* Probability of looking at the target (rather than the distractor) over time. The solid horizontal bar depicts the time region where infants were more likely to look at the target when it was unrelated to the prime ( $p < .05$ , using parametric cluster test, see text).

Based on the prior findings from the word comprehension task (Mani, 2010), we expected that phono-semantic priming would result in reduced looks to the target in the related-prime condition. However, because of the open-ended nature of our task, we did not have an a priori hypothesis about when this priming effect would emerge. Since our participants were not directed to look at either of the images, their fixation patterns cannot be predicted on the basis of the word comprehension studies. Instead their fixations reflect processes that occur spontaneously during passive viewing (e.g., object recognition or spontaneous verbal labeling). These processes could be synchronized to the onset of the visual stimuli in multiple ways. Thus, we needed a statistical approach which would allow us to compare looking patterns in the related-prime and unrelated-prime conditions across the entire time period when the images were present.



We adopted a strategy for identifying time windows with significant priming effects that allowed for flexibility while maintaining the probability of Type I error at  $p < .05$ . Specifically, we used a non-parametric cluster-size statistical test following Maris and Oostenveld (2007), which permits testing each time point without inflating the likelihood of generating a false positive. For each time point, we performed a logistic mixed-effects regression on fixations to the target, with prime condition as a fixed effect and subjects and items as random effects, implemented using the lme4 library (Bates, & Maechler, 2010) in R, version 2.11.1 (R Development Core Team, 2010). Time points were grouped into larger windows by identifying clusters of adjacent time points with significant prime condition effects ( $p < .05$ , two-tailed), and the  $z$  statistic for each time point within the cluster was summed to yield a summary statistic for the cluster. To determine the probability of observing a cluster of that size by chance, we conducted 1000 simulations where the condition labels for each trial were randomly assigned. For each of the 1000 sets of re-shuffled data, we implemented the analysis described above and saved the summary statistic for the largest cluster that was identified. A cluster of time points from the original data was considered to show a significant effect of prime condition if its summed  $z$  score was greater than the summary statistic of the largest cluster found in 95% of the re-shuffled simulations. Thus the total probability of a false positive is kept below .05. The  $p$ -values reported here reflect the proportion of re-shuffled simulations that found clusters with summary statistics as large or larger than the reported cluster.

Using this technique, we found fewer looks to the target in the related-prime condition from 550-850ms following target-distractor picture onset ( $p < .05$ ). To test for weaker but long-lasting priming effects, we repeated the analysis described above with a

relaxed alpha of .2 for each time point, while maintaining an overall alpha of .05 for each cluster. In this analysis, we observed an effect of prime condition in the 517-1383ms time window ( $p < .01$ ), again reflecting an inhibitory priming effect, consistent with the phono-semantic priming results from Mani (2010).

## **Discussion**

This study explored whether infants engage in spontaneously verbal encoding in a non-communicative and non-linguistic task. Infants saw a stream of pictures, none of which were labeled. Nevertheless, their looking patterns on the test trials were influenced by the names of the objects that they had previously seen. Specifically, when the label of the prime object was phonologically related to a close semantic associate of the target, infants were less likely to look at this object. Since phono-semantic priming depends upon the phonological form of the prime, this result indicates that infants internally generated the labels for the visually-presented images. Thus, early in development, language is active, even outside of communicative contexts or linguistic tasks, suggesting it may play a role in internal thought.

These results support the hypothesis we derived from the psycholinguistic model: since internal verbalization involves a subset of the processes involved in language production, it should emerge early in language acquisition. Our findings also provide additional evidence that the 24-month-old infants have a lexicon with the architectural features required to support phono-semantic priming (e.g., phonological neighborhood structure and a close coupling of phonological and semantic representations), confirming the findings from word-recognition tasks (Mani et al., 2012; Mani, 2010).

There are several possible paths from phono-semantic priming to the observed decrease in target looking in this task. It is worth noting that studies using a similar paradigm by Mani and colleagues have also found a decrease in looking to targets following phonologically and phono-semantically related primes (Mani, 2010; Mani & Plunkett, 2011; c.f. Mani et al., 2012). In particular, they find a decrease in target looking in studies with 24-month-old infants where there is only a single phoneme overlap between the prime and subprime or target, which matches the conditions of the current study. In their studies, Mani & Plunkett (2011) and Mani (2010) suggest that the phonological activation of the prime label ultimately inhibits activation of the target label, delaying word recognition. However, it is unclear that inhibition of the target label should necessarily lead to a decrease in looking in the current study, which does not involve word recognition. Another possibility is that accessing the prime label actually facilitated processing of the target object, such that infants were able to retrieve the label more quickly and look away sooner. This explanation is consistent with fixation patterns in Zelinsky & Murphy's (2000) recognition memory task, where adults looked less at objects which required less time to verbalize.

While these findings challenge Vygotsky's specific claims about the age at which internal speech emerges (1934/1987), they lend support to his broad proposal that external symbols play a central role in internal mental life. Researchers in the Vygotskian tradition have argued that internal verbal representations emerge as private speech declines between the ages of 4 and 7 (Winsler, 2009). There are two ways in which this theory could be modified to account for our findings. First, one could retain the premise that private speech is a gateway into internal speech and draw a theoretical distinction

between spontaneous verbal activity, which is present in infancy, and true internal speech. This would require further development of the theory and a re-examination of prior research which has used phonological representation as an index of internal speech (Al-Namlah, Fernyhough, & Mein, 2006). Second, one could seek another explanation for the disappearance of private speech such as the acquisition of taboos about talking to oneself (Duncan & Tarulli, 2009).

Our data raise new questions about how verbal encoding changes during development. In adults, the phonological encoding of pictorial stimuli is well-established in short-term memory tasks and language production tasks, but the results from nonlinguistic visual-search tasks are inconsistent. For example, Meyer et al. (2007) found that adults looked longer at distractors when they were homophonous with the target, suggesting that participants were accessing the image labels. On the other hand, Telling (2008) failed to find any effects of partial phonological overlap between target and distractor labels in a range of similar tasks, suggesting that adults may not be activating phonological labels during visual search. In another visual search experiment, Zelinsky and Murphy (2000) did not find effects of verbal activation in adult participants. In contrast, our results demonstrate that infants implicitly name the objects that they see in a nonlinguistic task with no memory component. We have considered three explanations for this difference.

First, the free-viewing task could be more sensitive to the effects of implicit verbal labeling than the visual-search task. While more research is needed to evaluate this possibility, our initial studies suggest that it is wrong. We have found that adults show the same pattern of effects in a free-viewing task as they do in the visual-search task: greater

looking to homophonous targets but not to phonologically-related words (Khan, Fitts, & Snedeker, in prep.).

Second, over development we may gain the ability to limit the activation of linguistic representations to contexts where they are useful (short-term memory tasks) or necessary (word production tasks). This shift could be part of a broader developmental change in the ability to use goals to control cognitive processes, linked to the maturation of prefrontal cortex and the development of executive functions (see e.g., Zelazo, Carlson, & Kesek, 2008).

Finally, perhaps adults, like infants, spontaneously activate linguistic labels of objects they see, but this activation is undetectable with current paradigms because the adult lexicon is so much larger than the child's. A typical 24-month-old produces about 300 words (Dale & Fenson, 1996), a typical adult knows about 60,000 (e.g., Miller, 1996; Nagy & Herman, 1987; Pinker, 1994). As a result, a given word will have many more phonological neighbors for an adult than for a 2-year-old, thus the priming effect for any given phonological neighbor may be diminished. This hypothesis could explain why adults in visual-search tasks show effects of homophony (Meyer et al., 2007) but not phonological overlap (Telling, 2008): homophones necessarily have a greater degree of phonological overlap than non-homophonous pairs.

The use of spontaneous verbal encoding in infancy is relevant to our broader understanding of the relationship between language and thought. The influence of language on thought is often studied by looking at how speakers of different languages perform on what appear to be non-linguistic cognitive tasks. When differences between populations are found they are often interpreted as evidence that language affects the concepts that are

available or salient to the speaker (see Gentner & Goldin-Meadow, 2003 for examples). Our findings suggest another possible explanation: linguistic representations may become active in many tasks which do not explicitly require language comprehension or production. When this happens, differences in the language itself may influence performance, even if speakers of both languages have access to the same conceptual representations. This account correctly predicts that verbal interference can disrupt cross-linguistic cognitive differences (see e.g., Winawer et al., 2007) and that a bilingual's pattern of performance will often depend on the language in which they are tested (Boroditsky, Ham, & Ramscar, 2002; Barner, Iganaki, & Li, 2009). Such verbal interference effects have not been explored by developmental psychologists, but the present results suggest that they may begin very early in life.

## Chapter 4. Implicit activation of verbal labels during visual object processing<sup>3</sup>

The connection between a familiar concept and its verbal label is often so engrained that going from the former to the latter feels automatic. The process of retrieving the word “apple” when talking about the familiar red fruit, for example, typically goes by effortlessly and unnoticed. The fluid transfer from concepts to words in speech production and from words to concepts in comprehension leads to the question of how these representations interact in the absence of communication. Does calling up a concept spontaneously lead to activating the corresponding word form?

The role of linguistic representations in our mental lives has been a topic of extensive debate. From Whorf (1956) to Chomsky (2007), there have been numerous theories placing language at the heart of cognition, certainly viewing it as more than a mere tool for communication. The specific proposals vary widely on the nature of the interaction between linguistic and nonlinguistic representations, but they share the notion that language is involved in internal thought. One starting point for investigating this issue is to explore what contexts elicit internal linguistic representations and what form these representations take. Introspection tells us that verbal representations are sometimes used in non-communicative contexts. Whether it is in mentally rehearsing to-be-remembered material or in the conscious experience of an inner speech stream, it is clear that we sometimes involve language in our inner thoughts. However, there are cases where introspection fails us, where we do not have clear access to the representations we rely on. For example, the nature of the representations underlying mental imagery was the topic of

---

<sup>3</sup> The study described in this chapter was conducted in collaboration with Whitney Fitts and Jesse Snedeker.

a longstanding debate (e.g., Kosslyn, 1994; Pylyshyn, 1973). So, the question remains: is activating linguistic representations in silent thinking a widespread phenomenon or something that only occurs during specific tasks? The current study starts to bridge this gap by investigating verbal activation during a simple visual object processing task. If verbal representations are recruited just from encoding a visual object, a process that occurs constantly in our daily lives, then it would appear that verbal representations are indeed ubiquitous in our internal thoughts.

There is some work that suggests verbal labels for visual images are spontaneously recruited during seemingly nonverbal tasks. Noizet and Pynte (1976) used a free viewing task, in which participants were simply told to look at each picture until they had recognized the drawing. They found that people spent longer looking at pictures with longer names than those with shorter names. However, it is unclear whether the participants interpreted the instruction to silently identify the objects as a request to implicitly label the images. A similar word length effect was observed by Zelinsky and Murphy (2000) in the context of a recognition memory task. Adults saw four pictures on a screen followed by a 2.5 second pause and then judged whether a probe image was one of the original four. During the encoding phase of this task, adults spent longer inspecting images with multisyllabic labels than those with monosyllabic labels, both in terms of initial gaze duration and in terms of total fixation time. Additionally, in a second experiment, Zelinsky and Murphy (2000) replicated this effect with novel face-name pairs, which participants had learned before beginning the recognition memory task. The novel pairings ruled out the possibility that fixation durations were actually determined by conceptual complexity or other possible confounding factors.



There is also some evidence of verbal encoding during visual search tasks. Meyer, Belke, Humphreys and Telling (2007) found increased visual attention to distracters that were homophonous with the target, even though verbal labels were not presented during the search task. The most obvious way of accounting for this result is via verbal encoding of the target image, since the targets and distracters were unrelated except for the shared label. However, other studies have failed to find verbalization effects in visual search. In a series of experiments, using the same visual search paradigm as Meyer and colleagues, Telling (2008) did not observe any competition from distractor images whose labels were phonologically similar to label of the target image, either in terms of onset overlap or rime overlap. Surprisingly, these effects were absent even when written words were presented rather than images, suggesting perhaps that participants developed specific encoding strategies over the course of the task. The absence of phonological encoding during visual search is also reported in Zelinsky and Murphy (2000). Participants who learned the face-name pairs (see above) also completed a visual search task with these items. However, the visual search task did not yield any name length effects, suggesting that participants were not labeling the faces as they searched for the target.

These apparent discrepancies can be reconciled if we assume that implicit labeling is a task-specific strategy that is used under a fairly limited set of circumstances. Short-term memory tasks, such as the recognition memory task used by Zelinsky and Murphy (2000), lend themselves to using phonological rehearsal, perhaps encouraging participants to strategically encode the pictures using verbal labels. Visual search, on the other hand, only requires holding a single item in memory and so does not necessitate such strategies. While intuitively attractive, this explanation is incomplete as it does not account for

increased attention towards homophonous distractors during visual search (Meyer et al., 2007). We can gain some traction on this issue by considering different levels of representation in lexical access, for example, distinguishing between word-form level representations and phoneme representations (e.g. Levelt, Roelofs, & Meyer, 1999). It is possible that the homophone effects reflect some level of lexical activation, in the absence of phonological (phoneme level) activation. In this case, the failure to find evidence for implicit phonological activation in visual search tasks (Telling, 2008; Zelinsky & Murphy, 2000) does not conflict with the homophone effect in a similar task. Taken together, these results suggest that implicit lexical activation occurs spontaneously across a range of visual tasks, including both memory and search tasks, but phonological access only occurs in tasks like short-term memory tasks, which prompt the use of verbal encoding strategies.

Recent results from infant studies, however, are at odds with this account of implicit verbalization. In a series of studies, Mani has found that 18- and 24-month-old infants activate the phonological label for objects that are visually presented (Mani, & Plunkett, 2010; 2011; Mani et al, 2012). On critical trials in these studies, a single prime image is displayed in silence; this is followed by a split screen of the target and distractor images, and then the target object is labeled. Infants' looking to the target is affected by the phonological relationship between the labels for the target and prime images, even though the prime label is never heard. This phonological priming is present even in a completely non-linguistic version of this task where neither the target images nor the primes images are labeled (Khan, Geojo, Wang, & Snedeker, submitted), indicating that 24-month old infants spontaneously recruit verbal representations in a non-communicative context..

There are two explanations for why phonological activation is absent in visual search tasks with adults but present in visual tasks with infants. First, adults might be less likely than infants to spontaneously activate linguistic representations in non-communicative contexts. This conclusion is at odds with other research that suggests an increase in the use of internal verbal representations through childhood; for example, preschool aged children are less likely than adults to rely on verbal representations in working memory tasks (for review see Hitch, & Halliday, 1983). However, it is possible that infants spontaneously engage in verbal encoding, whereas adults only recruit linguistic representations in contexts where it is strategic to do so.

A second possibility is that the paradigm that was used in the infant studies is more sensitive to phonological activation than the method employed in the visual search studies. The preferential-looking paradigm used by Mani and colleagues presented the prime image in isolation, so the infant only attended to a single item at the point of implicit labeling. In contrast, in the visual search tasks, adults would need to access the labels of four images presented simultaneously for verbalization effects to be detected. Implicit verbalization effects in visual search tasks are investigated by measuring the fixation time to the non-target distractors (e.g., the homophone distractors in Meyer, et al., 2007). Thus, the chances of finding verbalization effects are obscured by all of the other factors that cause the participants to move their eyes away from a distractor image, such as color or shape. For example, in Zelinsky and Murphy's (2000) visual search paradigm, a name length effect might not be detected in participants' eye movements as they might move their eyes away from a long-name distractor or linger longer on a short-name distractor based on visual features. In contrast, in the priming paradigm, implicit verbalization is probed by

measuring looking times to the target image, so the task demands do not draw eye movements away in a manner that reduces sensitivity to verbalization effects.

The current study explores implicit verbal activation in adults using a paradigm that allows direct comparison to the infant results. Using a trial structure based on the studies by Mani and colleagues, we examined priming based on the labels of images presented in silence. To the extent that such priming is observed, we can conclude that adults spontaneously accessed these verbal representations. We included both homophone and phonological-onset overlap priming in this study. As we noted above, the presence of homophone effects in visual search tasks, and the absence of phonological effects, suggests that lexical (word-form) representations might be recruited more promiscuously than phonological ones. This hypothesis suggests that a similar discrepancy might be present in a viewing task as well. In contrast, if adults perform this task in the same way that infants do (and activate phonological representations) then we should see phonological-onset priming, as well as homophone priming in our adult participants.

## **Experiment 1**

The first experiment in this study adapted Mani and Plunkett's (2010) preferential looking paradigm to study implicit label activation in adults. By keeping the procedure similar to the infant studies, which have found evidence of phonological activation during picture viewing, we can investigate whether the absence of such findings in adults should be attributed to developmental change or to differences across tasks. While the basic trial structure used in this experiment paralleled that of the infant studies, there were some necessary changes. First, a memory component was added to the design so that adults had some motivation to attend to the images. The number of to-be-remembered images (84

images per block) exceeded the bounds of working memory, so verbal rehearsal could not be an effective memorization strategy and therefore should not lead to phonological activation beyond the spontaneous initial encoding of the items. Second, in addition to prime-target pairs that shared their phonological onset (*cherry, chair*), participants also saw three other types of prime-target pairs: homophones (*bat*-baseball instrument, *bat*-animal), semantically-related pairs (*chicken, egg*) and phono-semantically related pairs (*log, key*). Including different trials types meant that there were fewer trials instantiating any single relationship, making it less likely that adults would become aware of the relevant patterns. Also, the homophone and semantically-related items allowed us to make comparisons to the visual search studies in adults which found effects of both of these types of relatedness (Meyer, et al., 2007). Finally, we increased the amount of phonological overlap in the phonological-onset and phono-semantic items. All phonologically related primes and targets, and primes and subprimes for the phono-semantic condition, overlapped in at least the first two phonemes, rather than just the initial segment. This was necessary due to the considerably larger size of adult vocabularies, compared to infant vocabularies. A single initial segment is likely to be shared across many more words for an adult and so, due to fan effects in spreading activation (Anderson, 1974), the priming effect for any one word would be considerably weaker. Increasing the amount of overlap increases the strength of the priming effect.

## **Methods**

### *Participants*

Thirty-two native English speakers between 18 and 30 years old (20 female, mean age 21 years) were recruited from the Harvard University subject pool. An additional two

participants completed the experiment but were not included due to excessive eye-tracking data loss (more than 50% unusable trials in at least one condition). Participants either received partial course-credit or \$10 as compensation.

### *Materials*

Each trial in this study consisted of a triplet of images: the prime, the target and the distractor. Eighty-four triplets were created in four different categories, which were defined based on the relationship between the prime and the target: 12 homophone triplets, 24 phonological onset triplets, 24 semantic triplets, and 24 phono-semantic triplets. For homophone triplets, the prime and target shared the same verbal label but were unrelated in meaning (e.g. *bat*-sports equipment and *bat*-animal). There were fewer items in the homophone category than in the other categories due to the difficulty of finding depictable homophone pairs with reliable name agreement. Phonological onset triplets contained prime and target images whose labels share the same initial sounds but are semantically unrelated (e.g. *cherry* and *chair*), while semantic primes and targets were related conceptually but not phonologically (e.g. *chicken* and *egg*). In phono-semantic triplets, the label for the prime image shared its phonological onset with an unseen intermediary (the subprime) which was semantically related to the target. For example, *log* would be used as a prime for *key*, with *lock* as the subprime, which was neither seen nor heard. In all cases, the distractor image was semantically and phonologically unrelated to both the prime and the target. Since prior studies have found a correlation between looking time and word-length (Zelinsky & Murphy, 2000), distractor images in each triplet were chosen such that their labels matched the target in terms of the number of syllables.

For each category, half the trials were presented in the Related Prime condition, in which the prime and target were related as described above. The other half were presented in the Unrelated Prime condition, in which the prime, target and distractor were all phonologically and semantically unrelated to each other. Unrelated Prime trials were created by shuffling the prime images, such that they were no longer presented in the same trial as the related targets. For example, in the Related Prime condition, *bacon-baby-igloo* and *penny-pencil-lighthouse* are presented as prime-target-distractor combinations, and in the Unrelated Prime condition these are shuffled into *penny-baby-igloo* and *bacon-pencil-lighthouse*. Target-distractor pairings were held constant across Related and Unrelated conditions. The experiment was split into three blocks of 28 trials, each including an equal number of Related Prime and Unrelated Prime trials from the different categories, presented in random order. Eight experimental lists were created such that, for each item, condition (Related Prime or Unrelated Prime), target position (left or right) and block order were counterbalanced across participants.

*Phonological relatedness assessment.* For the phonological onset prime-target pairs and the phono-semantic primes and subprimes, shared phonological onset was defined as overlap of the first consonant-vowel or vowel-consonant sounds of the word. Intuitions about initial phonological overlap were confirmed using two measures. First, IPA transcriptions from Webster's New World College Dictionary of American English indicated identical initial phonemes. Second, in a brief survey, at least 8 out of 10 American English speaking respondents agreed that the initial consonant-vowel or vowel-consonant sounds were the same across both words.

*Semantic relatedness assessment.* Surveys posted on Amazon Mechanical Turk were used to assess the semantic relatedness of the prime-target, prime-distractor, target-distractor, and the subprime-target pairings for both the Related Prime and Unrelated Prime conditions. On each survey, respondents were asked to rate item pairs, presented as written words, on a scale from 1 to 7 (semantically unrelated to extremely semantically related). In addition to the test items, each survey included filler items that were either strongly related or completely unrelated to encourage participants to use both ends of the scale. Since the shared name affected participants similarity judgments, homophone prime-target pairs were assessed in separate surveys using polysemes and other homophones as fillers. Each pair was rated by 10 respondents. We used an average rating of 5 as the criterion for a pair to be semantically related, with the exception of window-door which was rated slightly below this cut-off but was still included in the final stimulus set. We confirmed that semantic prime-target pairs and phono-semantic subprime-target pairs in the Related Prime condition were judged to be semantically related ( $M=6.18$ ), and all other pairs were not ( $M=1.68$ , maximum=3.6). Further, apart from semantic pairs, primes and targets were comparably related in the Related Prime and Unrelated Prime conditions ( $t=0.08$ ,  $p=0.940$ ).

*Image selection.* Colored line-drawings of familiar objects were selected from two sources: 140 from Rossion and Pourtois's Snodgrass and Vanderwart 'Like' Objects (2004) and 112 from Google Image searches, with the latter edited to look stylistically similar to the former. All images were cropped and resized to 267x200 pixels, so that the images covered approximately the same amount of space on the screen. Email surveys were used to ensure naming consistency for all images that were selected. At least 10 participants

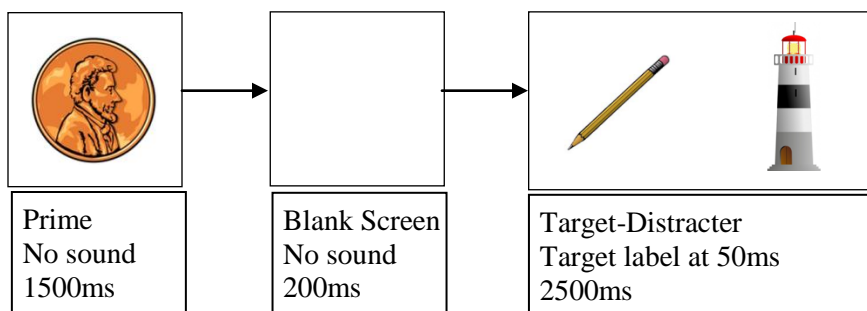


named each image. Other than images with homophonous labels and images in the semantically-related trials, we used an inclusion criterion of 80% naming consistency. Different inclusion standards were used for homophone items due to people's tendency to spontaneously produce adjectives when labeling these images e.g. fish tank rather than tank). For these items, prime images had to be named with at least 70% consistency, and target images were selected such that the intended word was usually part of the produced label (M=87%, minimum=64%). We relaxed the naming consistency requirements for the semantic primes and target to allow for near-synonyms, such as chicken and hen. For these items, that is the images with two common labels, both labels were tested for semantic relatedness in the procedure described in the earlier section. In general, for prime-target pairs with asymmetric naming consistency, the image that elicited more consistent labeling was used as the prime. Since we are probing implicit activation of the prime label, it is critical that this image elicits the intended label; if it does not, then the intended priming pattern will not occur even if implicit labeling does occur. Further, since participants heard the target label explicitly, for the target image, image recognition was more critical than the particular label that spontaneously came to mind.

### *Procedure*

The experimental procedure, including stimulus presentation and eye tracking recording, was implemented using E-Prime software and the session was conducted using a Tobii T60 remote eye-tracker. At the beginning of the session, participants were told that they would be shown a series of images which they should try to remember because they would be tested on them later.

Each trial began with fixation at center of the screen. After the participant fixated this location, the prime image appeared in the center of the screen and remained on-screen for 1500ms. This was followed by a blank screen for 200ms, and then the target and distractor images appeared on the two sides of the screen for 2500ms. Fifty milliseconds after these images appeared, the label for target image was played over the speakers. A sample trial progression is depicted in Figure 5.



*Figure 5.* Schematic of a trial in related-prime phonological onset overlap condition (prime *penny*, and target *pencil*). In the unrelated-prime condition, the sequence remains the same except for the prime image, which would be replaced by *bacon*, with is the related-prime for a different target.

The experiment was divided into three blocks of 28 trials, with 10 memory test trials at the end of each block. Each memory test trial consisted of a single image for which the participant gave an “old” or “new” judgment. The five old items were always exactly the same images as the ones presented earlier. The five new items were completely different items from ones that had been seen, rather than different instances of the same kinds (e.g. two different images corresponding to the label penny).

After all of the experimental trials, each participant was given a post-test questionnaire to determine if they had noticed the various prime-target relationships present in the stimuli. They were asked 1) what they thought the experiment was about,

and whether they noticed anything about 2) the pictures, 3) the words, 4) the relationship between the two pictures and the picture that preceded them on each trial. The questions became progressively more specific to give participants the greatest opportunity to indicate whether they noticed the prime-target relationships. We coded the responses in terms of whether or not participants noticed relationships between the image labels. All participants noticed the presence of homophones and 22 out of 32 were aware of other types of phonological relatedness as well.

Following the post-test questionnaire, the participants were shown all 252 images individually on the screen and asked to name them. This allowed us to confirm that participants labeled the images with the expected nouns.

#### *Data processing*

*Trackloss.* Eye tracking data was recorded at 60Hz and each sample included a validity rating for the recording from each eye. Only samples that reported valid recording from both eyes were included in data analysis. Further, we excluded trials where fewer than 50% of the expected number of samples met this criterion; 4% of trials were excluded for this reason.

*Fixation coding.* The duration of the target-distractor screen for each trial was divided into 25 hundred millisecond time-bins. For every 100ms, we coded whether there were any fixations in the 267x200 pixel region of the target image or the distractor image. Since we were interested in the amount of target looking relative to the distractor, time bins with no fixations in either region were excluded from further analysis. Thus, for each 100ms bin where either the target or distractor was fixated, we analyzed whether the target was fixated.

## Results and Discussion

### *Memory accuracy and name agreement*

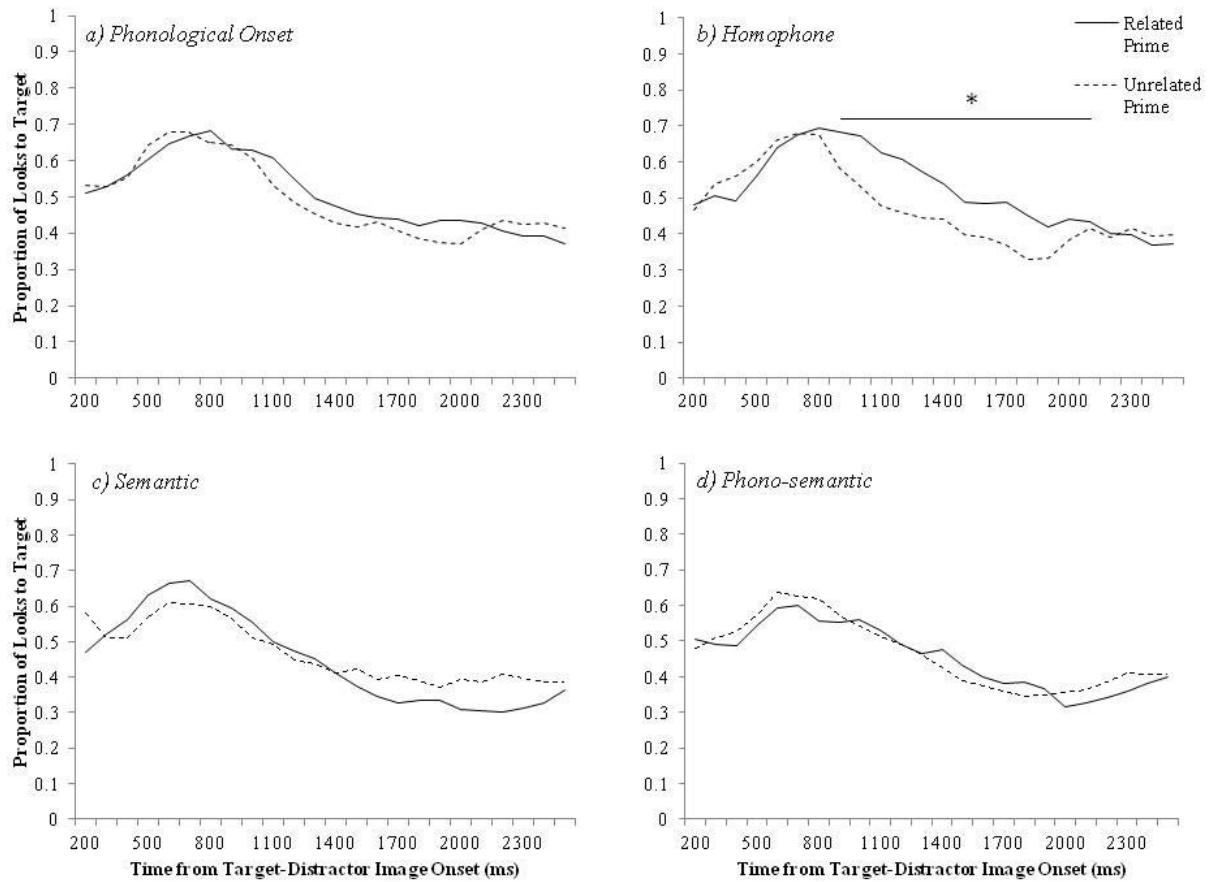
Participants were generally successful at recognizing old and new images during the memory components of the task ( $M=89\%$ ,  $SD=9\%$ ). Also, the naming post-test confirmed that the images used in the experiment generally elicited the expected labels for the primes ( $M=91\%$ ), targets (95%) and distracters (90%). Naming accuracy was likely higher for target images as participants heard labels for these images during the main experiment. Given the generally high rate of expected naming, we present analyses including all trials regardless of whether the image was correctly named in the post-test.

### *Eye tracking results*

We did not have a priori hypotheses about the specific timing of each priming effect, so we used a data-driven technique for finding time windows with significant differences in target fixation probability in the Related Prime and Unrelated Prime conditions. A non-parametric permutation test, adapted from (Maris and Oostenveld, 2007), was used to allow for flexibility while maintaining the probability of Type I error at  $p<0.05$ . For each 100ms time bin, we performed a mixed-effects logistic regression with Prime Relatedness as a fixed effect, and with random intercepts for subjects and items; the different trial categories (homophone, phonological onset, phonosemantic, semantic) were analyzed separately. Next, we grouped adjacent time bins where there was a significant effect of relatedness into clusters. The  $z$  statistic of each time bin within a cluster was summed to yield a summary statistic for that cluster. A liberal criterion,  $p<0.25$ , was used for a time bin to be included in a cluster so that we could detect small but long-lasting priming effects. It is important to note, however, that the criterion for including a time bin in a cluster is

orthogonal to the process by which cluster significance is tested, so it does not affect the likelihood of a false positive. To determine the probability of observing a cluster of that size by chance, we conducted 1000 simulations where the condition labels (Related Prime or Unrelated Prime) for each trial were randomly assigned. The cluster-finding algorithm described above was then implemented on each set of reshuffled data. A time bin cluster was considered to show a significant effect of Prime Relatedness if its summed  $z$  statistic was greater than the summed  $z$  statistic of the largest cluster found in 95% of the reshuffled simulations. Thus, the total probability of a false positive is kept below .05. The  $p$ -values reported here reflect the proportion of re-shuffled simulations that found clusters with summary statistics as large or larger than the reported cluster.

Using this approach, we found a significant priming effect for the Homophone trials, but we did not observe significant priming for the Phonological-Onset, Phono-Semantic or Semantic trials (Figure 6). When the target was related to the prime via homophony, participants were more likely to fixate the target image, compared to unrelated control trials, from 900-2100ms following the target-distractor image onset ( $p < 0.01$ ). While we did not observe a significant time window for the effect of semantic-relatedness emerged later, we did detect two adjacent clusters of time bins with significant differences, with fewer looks to the target image following a related prime, from 1500-1900ms window and 2000-2500ms ( $p = 0.31$  and  $p = 0.12$ , respectively). Thus, there was a marginal effect of semantic priming, since the result would have reached significance if the single time bin from 1900-2000ms ( $p = 0.29$ ) had shown a larger effect; if the  $p$ -value for this time bin had been less than .25 then the two clusters mentioned above would have been combined into a single, more robust cluster.



**Figure 6.** Experiment 1 eye tracking results. Proportion of looks to target image for a) Phonological-Onset, b) Homophone, c) Semantic, and d) Phono-semantic relatedness trial types.

These results parallel the findings from visual search tasks (Meyer, et al., 2007; Telling, 2008), in which homophony between the target image and one of the distractors affects eye movement patterns, but partial phonological overlap does not. On the other hand, the absence of phonological-onset and phono-semantic effects in the current experiment stands in contrast to the results from 18 and 24 month olds in a similar task (Mani & Plunkett, 2010; 2011; Mani et al, 2012). Overall, the similarity of the present findings to those from other adult implicit labeling paradigms suggests that adults do not spontaneously activate phonological representations during visual object processing. Since the current experiment uses a paradigm that parallels the infant studies, it seems that

differences in testing procedure cannot account for why phonological-onset and phonosemantic priming is observed in infant implicit naming studies, but are absent in studies with adults; a developmental explanation is needed.

The divergence of the Homophone and Phonological-Onset conditions, which is also observed in the visual search paradigm, is open to several interpretations. One possibility is that homophony and phonological overlap effects are on the same continuum, but only the homophone effects are detectable because total overlap leads to greater priming. Another possibility is that there is a qualitative difference between the homophones and phonological-onset pairs that affects priming in this task. In particular, differences could arise if participants do not activate phonological representations of the image labels (i.e. phonemes), but do activate word-form representations (i.e. lexemes). For example, an image of a baseball bat could prime *bat*-animal by activating the word-form *bat*, but it could only prime *bag* by activating the phonemes that the two words have in common.

We further examined the Phonological-Onset trials to determine whether the results suggest that a smaller version of the Homophone priming effect might be present. While there were no time windows with significant priming effects for the Phonological-Onset condition, there were some interesting trends. First, the two largest non-significant clusters, 1100-1500ms ( $p=0.31$ ) and 1900-2100ms ( $p=0.68$ ) overlapped with the time window indicated in the Homophone analysis, and in both cases, there were more looks to the target in the Related Prime condition. Second, we estimated our power to detect an effect of phonological priming, assuming that it would occur in the same time window as the homophone priming effect and would be approximately half the size. Our observed power for a t-test, by subjects, in this window was 0.7 (Lenth, 2006-9). Third, we compared

average proportion of target looking in the 900-2100ms window in related and unrelated trials for the Homophone and Phonological-Onset conditions, using a mixed-effects regression, with the maximally appropriate random-effects structure that converged, for subjects and items, and only found a marginally significant interaction ( $\beta=0.07$ ,  $SE=0.04$ ,  $p=0.06$ ). Together, these features of the data cast some doubt on the absence of phonological-onset priming in this experiment. Experiment 2 focuses on phonological-onset priming and addresses these concerns.

## **Experiment 2**

In Experiment 2, we explored whether this paradigm is in fact sensitive to phonological onset priming in adults. To this end, we changed the design from Experiment 1 in two ways. First, we doubled the number of phonological onset trials so that we would have greater statistical power to detect any small effects that are present. Second, we included another condition, between-subjects, where the prime is auditorily labeled. This allowed us to determine whether we could detect phonological priming effects in this paradigm, with these items, when we ensure that the prime label was activated. Observing priming effects in this condition indicates that our paradigm is sensitive to phonological activation when labels are explicitly presented and suggests that the lack of priming in the other condition is due to an absence of phonological activation when objects are viewed in silence.

## **Methods**

### *Participants*

Sixty-four native English speakers between 18 and 28 years old (41 female, mean age 21 years) were recruited from the Harvard University subject pool. An additional eight



participants completed the experiment but were not included due to excessive eye-tracking data loss (more than 50% unusable trials in at least one condition). Participants either received partial course-credit or \$10 as compensation.

### *Materials*

There were 108 trials in this experiment: 48 phonological onset trials (Related Prime or Unrelated Prime) and 60 filler trials. The prime-target-distractor triplets from Experiment 1 were mostly used again in Experiment 2, with a few changes. In addition, the phono-semantic triplets were converted to phonological onset triplets by replacing the target image with the subprime. For example, the phono-semantic prime-target pair *log-key* from Experiment 1 was converted to *log-lock* in Experiment 2. These new target images were normed for naming consistency and semantic-relatedness using the same procedures and criteria as Experiment 1. Additionally, we wanted to increase the number of trials with no phonological relatedness to counteract the increased number of phonological onset triplets (24 in Experiment 1, 48 in Experiment 2). These filler trials came from three sources. Twenty-four new filler trials were added to the design, with primes, targets and distracters all unrelated to each other. Further, all 12 homophone trials were presented in the Unrelated Prime condition and treated as fillers. Finally, like in Experiment 1, there were 24 semantic trials, with half presented in the Related Prime condition and half presented in the Unrelated Prime condition. The Related semantic trials were included because participants were typically aware of semantic relatedness and we hoped that this would prevent them from realizing that phonological relatedness was of particular interest in the study.

As in Experiment 1, eight experimental lists were created such that Prime Relatedness, target side and block order for each phonological onset item was counterbalanced across participants. Each block in Experiment 2 consisted of 36 trials.

### *Procedure*

Participants were randomly assigned to the Prime-Labeled or Prime-Unlabeled condition. The Prime-Unlabeled condition was identical to Experiment 1 except for the differences in stimuli detailed above. In the Prime-Labeled condition, the procedure was modified by adding a recording of the label for the prime image that began when the prime image appeared on the screen. Thus, in the Prime-Labeled condition both the target and the prime were explicitly labeled, obviating the need for implicit labeling. In contrast, in the Prime-Unlabeled condition, like Experiment 1, only the target was auditorily labeled, so phonological priming effects would depend on implicitly labeling the prime image.

As before, participants completed a post-test questionnaire to assess whether they noticed the presence of phonologically related items. Out of 64 participants, 44 mentioned phonological relatedness in their answers, and this did not differ based on whether they were in the Prime-Labeled or Prime-Unlabeled condition. Finally, participants were shown all of the images from the experiment individually and were asked to name them so that we could confirm naming consistency for the critical items.

### *Data Processing*

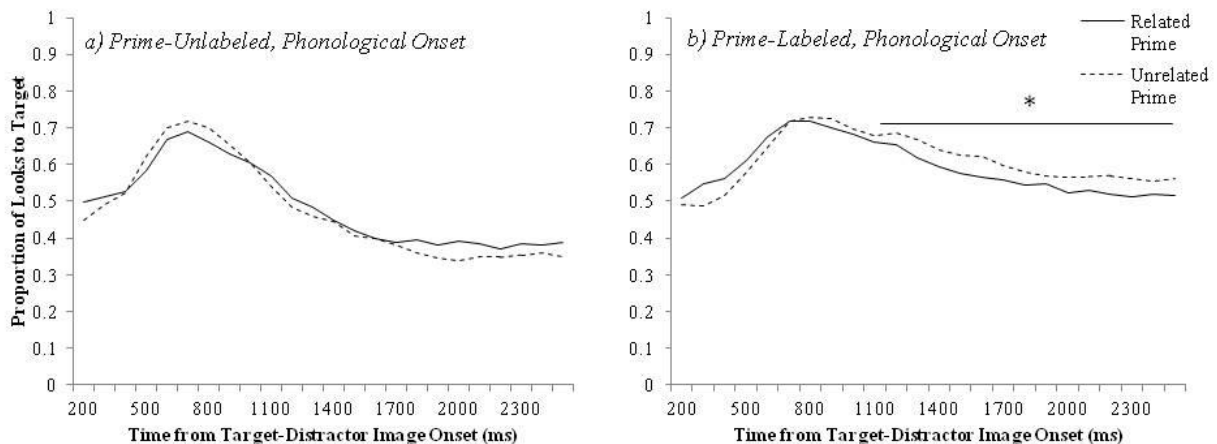
Eye tracking data was processed in the same manner as Experiment 1. 7% of trials were excluded from data analysis due to excessive trackloss, based on the criteria outlined earlier.

## Results and Discussion

### *Memory accuracy and name agreement*

Average accuracy on the memory trials in the Prime-Unlabeled condition was 85% ( $SD=10\%$ ), compared to 80% ( $SD=9\%$ ) in the Prime-Labeled condition ( $t(62)=2.06$ ,  $p<0.05$ ). We assessed naming accuracy for the critical trial images across these two conditions from the picture naming posttest. Overall, participants generally named pictures as expected in both the Prime-Unlabeled and Prime-Labeled conditions (across items,  $M=94\%$  and  $95\%$ , respectively). As expected, participants in the Prime-Labeled condition correctly named the prime images at a higher rate than those in the Prime-Unlabeled, since they had heard labels for these images in the main experiment (across items,  $95\%$  vs.  $91\%$ ,  $t(47)=3.80$ ,  $p<0.01$ ).

### *Eye tracking results*



*Figure 7.* Experiment 2 eye tracking results. Proportion of looks to target image following a Phonological-Onset prime image that was a) Unlabeled or b) Labeled.

We analyzed the eye tracking data separately for the Prime-Labeled and Prime-Unlabeled condition, using the cluster-finding procedure to find time regions with

significant priming effects (Figure 7). In the Prime-Unlabeled condition, there were no reliable differences between proportion of target fixations when the target and prime were phonologically related or unrelated ( $p > 0.3$  for the largest observed cluster). On the other hand, in the Prime-Labeled condition, there was a long-lasting phonological priming effect ( $p < 0.01$ ), emerging at 1100ms and continuing until 2500ms, the end of the trial. During this time, participants were less likely to look at the target image if it was phonologically related to the preceding prime image; this effect is in the opposite direction from the homophone effect in Experiment 1. We compared the average proportion of target fixations in this time window across the Prime-Labeled and Prime-Unlabeled conditions using a linear mixed-effects regression, with random intercepts for subjects and items. There was a significant interaction between labeling and relatedness ( $\beta = 0.07$ ,  $SE = 0.02$ ,  $p < 0.01$ ), further supporting the conclusion that phonological priming was only present when the prime was explicitly labeled.

The results from the Prime-Unlabeled condition strengthen the null effect of phonological priming from Experiment 1. In both experiments, simply viewing an image did not lead to phonological priming from the label of that image. This cannot simply be a failure of the priming paradigm used here, as the presence of phonological priming in the Prime-Labeled condition indicates that this paradigm is sensitive to phonological activation. This suggests that adults, unlike infants, do not spontaneously activate the phonological representation of the image label in noncommunicative contexts. We cannot fully rule out the possibility that implicit phonological priming does occur but was too weak to be detected in our task. However, we did not find an increase in the cluster size or reliability after doubling the number of items with phonological-onset overlap from

Experiment 1 to Experiment 2, as we might have expected if there was a real but weak effect. In addition, we conducted a power analysis, parallel to the one in Experiment 1, to estimate our power to detect an effect of phonological priming in Experiment 2. We assumed that the effect would occur in the same time window as the homophone priming effect in Experiment 1 and would be approximately half the size. Our observed power for a t-test, by subjects, in this window was 0.8 (Lenth, 2006-9).

## **General Discussion**

The experiments presented here demonstrated that viewing an image, without hearing or speaking its label, affects subsequent looking time to images that are related via homophony but not images that are related via partial phonological label overlap. This data pattern simultaneously suggests both the presence and absence of implicit linguistic activation during visual object processing, perhaps reflecting different levels of representation. The null effect of partial phonological overlap and the positive finding for homophone priming echo the results of the visual search tasks that have investigated implicit label activation (Meyer, et al, 2007; Telling, 2008; Zelinsky, & Murphy, 2000). On the other hand, the absence of phono-semantic and phonological overlap effects in the current study contrasts with the infant studies that have found such effects.

Given the results of these experiments, combined with the prior implicit labeling results with both adults and infants, our account of verbal activation during image recognition needs to address two puzzles in the data pattern: 1) Why do we find homophone priming but not phonological priming in adult implicit labeling experiments? 2) Why do infants, but not adults, show phonological priming in this paradigm? Below we consider three hypotheses.

One possibility is that adults do not engage in implicit labeling unless it is strategic for the task at hand (Zelinsky, & Murphy, 2000). Under this hypothesis, the absence of phonological effects in the current study and in the visual search paradigms is due to the lack of a working memory component in these tasks. The contrast with infants can be accounted for by suggesting that at the early stages of language acquisition, language activation might not be specialized or strategic. Alternatively, it is conceivable that image viewing might be a context where activating words is functional for an infant, who is engaged in word learning every day. It is more difficult under this story to explain the homophone effects, since the presence of homophone priming suggests that adults are also activating words in these tasks that lack a strategic verbal component.

Another possibility is that the phenomena are qualitatively similar but quantitatively distinct: spontaneous implicit labeling occurs in both adults and children and phonological representations are activated in both groups. However, in adults phonological priming is too weak to detect when only a few phonemes are shared. In adults, complete phonological overlap increases the strength of priming so that is detectable for homophones. Phonological priming is detectable in infants because the size of the lexicon is much smaller in infants, estimated at about 300 words for a typical 24-month-old (Dale & Fenson, 1996), compared to 60,000 for a typical adult (e.g., Miller, 1996; Nagy & Herman, 1987; Pinker, 1994). Thus, the fan size (Anderson, 1974) is simply much smaller for priming in the infant's lexical network. Consequently, the activation of one or two phonemes spreads to only a handful of known words resulting in detectable priming of the target.

There are some observations, however, that argue against the idea that phonological priming did occur in the current experiments but was too weak to be detected. For one, we did not observe phonological onset priming even after doubling the number of items in Experiment 2 in an effort to increase the sensitivity of the paradigm to such effects. Nor did we find any increase in the reliability of the non-significant phonological priming effects between Experiment 1 and Experiment 2. In addition, the phonological priming effect in the Prime-Labeled condition in Experiment 2 suggests that the current paradigm was sensitive to phonological activation. The repeated failure to find phonological overlap effects or word length in visual search tasks (Telling, 2008; Zelinsky & Murphy, 2000) also speaks against the idea that implicit activation occurs during visual object processing in adults.

Finally, we consider the possibility that adults spontaneously recruit verbal labels during visual tasks but not at a phonological level of representation. For example, visual object processing might result in activating a lexical item without activating all of the component phonemes of that word. Under this hypothesis, homophone priming effects occur in implicit verbalization studies because homophones share word-level representations (lexemes, e.g., Levelt, et al., 1999; but see Caramazza, Bi, Costa, & Miozzo, 2004). Phonological onset pairs, on the other hand, only share phonological level representations. Therefore, if visual processing spontaneously leads to lexical but not phonological activation, phonological priming does not occur in these tasks.

Under the hypothesis above, a strategic or task-dependent aspect to implicit language use is still necessary for explaining why different levels of verbal activation might occur across tasks and age groups. For example, phonological representations are recruited

to support working memory in visual tasks, as demonstrated by the word length effect during image encoding in Zelinsky and Murphy's (2000) recognition memory task. Experience or proficiency with the task at hand might also play a role in the form of implicit language use. Research in the domain of private speech (overt verbalization that is self-directed) has found that, while this phenomenon primarily occurs in young children, adults also engage in private speech when engaged in difficult or unfamiliar tasks (Sanchez Medina, Rubio, & De la Mata Benitez, 2009). In fact, contrary to the traditional Vygotskian model (Vygotsky, 1934/1987), these authors suggest that the transition from overt private speech to inner speech is a function of task-specific factors, such as proficiency and experience, rather than age or developmental stage. This explanation can be extended, then, to include a continuum of levels of representation, rather than just the binary distinction of inner speech versus private speech. There are many different levels of representation for a single word, from central conceptual representations to peripheral sub-articulatory representations. Lexemes and phonemes are just two of the levels of representation on this continuum. The proposal here is that 1) inner speech may vary in terms of which of these levels of representation it recruits, and 2) the form of inner speech depends of task proficiency, with more specified representations being recruited in less familiar or more difficult tasks.

The investigation of when visual object recognition leads to the activation of verbal representations forms one part of understanding when and how and why language enters internal thought. The homophone priming effect in Experiment 1 suggests that, in visual tasks, adults spontaneously activate some verbal representations of the image labels. However, this activation does not result in phonological priming. These results parallel



previous studies on spontaneous label activation in adults: visual search tasks which find homophone priming but not phonological priming (Meyer, et al., 2007; Telling, 2008; Zelinsky & Murphy, 2000). This consistent pattern suggests that word-level representations and phonological representations might be dissociable in implicit verbal label activation in adults, and that only the former are spontaneously activated during visual object processing.

## Chapter 5. Conclusions

The three studies in this dissertation have painted a heterogeneous picture of inner speech, as seen through the lens of implicit verbal labeling. Paradoxically, the results seem to simultaneously point to inner speech becoming more and becoming less prevalent through development. In this chapter, I discuss the results from these studies with respect to each other to try to unravel this contradiction. Through considering the task contexts in which verbal activation is observed in different age groups, a more nuanced view emerges of inner speech and its role in cognition.

To begin, the results and implications of each study are summarized here. In Chapter 2, we saw that adults and school-aged children spontaneously use a verbal encoding strategy when asked to remember images in a short-term memory task, whereas preschoolers relied solely on visual encoding. One interpretation of these results is that implicit verbal activation is automatic but late emerging, since it was observed in older children and adults but not in younger children. Alternatively, the data are also consistent with a picture of implicit verbal activation where it is task-related in a way that interacts with age, since the older age groups might understand the task of memory encoding differently than the 4-year-olds.

In Chapter 3, we saw that 24-month-old infants who are viewing images in a non-communicative context spontaneously implicitly activate verbal labels. We observed phono-semantic priming stemming from an image that was presented silently, indicating that the infants activated the phonological form of the word. This phonological activation spread through the lexical network, through semantic connections, as activation from an explicitly spoken word would. The infants' task in this study was simply to look at the

images and yet we found evidence of phonological label activation, suggesting that this process is spontaneous, rather than task-driven, in 24-month-olds.

The adult state for implicit verbal activation was explored in Chapter 4. Across two experiments, there was no evidence for phonological activation of verbal labels during picture viewing. This null effect contrasts the results from 24-month-olds in a picture viewing task, and is also unlike the results from adults in a short term memory task. However, the data pattern is not consistent with a complete lack of verbal activation, either, since we observed homophone priming stemming from an image that was presented silently. Since homophone image pairs are only related by virtue of their labels, such priming effects indicate that some level of lexical activation occurred. These results suggest that perhaps adults were recruiting abstract lexical representations, such as word-forms, without activating the corresponding phonological representations.

How do we reconcile the apparent increase in implicit language use through development observed in Chapter 2, with the apparent decrease that is seen in Chapters 3 and 4? Even with the studies with adults, there appears to be inconsistency, with robust phonological activation present in the experiment in Chapter 2 but absent in the experiments in Chapter 4. The story of implicit verbal activation – of inner speech – that emerges must consider both task and age, as well as their interactions. Although there are surely many other options that are compatible with the data at hand, the following paragraphs focus on just two of the possible developmental trajectories.

One possibility is that implicit verbal activation occurs automatically in infants and gradually becomes more task-constrained through development, such that verbal activation is limited to contexts where it is functional in adulthood. On this story, the

phonological form of “apple” might enter an infant’s mind any time he or she recognizes an apple in the environment, whereas adults will strategically recruit representations based on the task at hand. In a working memory task, for example, it might be useful to encode phonological distinctiveness and so the form “apple” would be activated. Other tasks might benefit more from, for instance, visual or conceptual distinctiveness, and so different representations would be recruited and relied on depending on the particular task context. For example, Konkle and colleagues found that conceptual distinctiveness was a key factor in performance on a visual long term memory test, but perceptual factors were more important when the same stimuli were used in a visual search context (Konkle, Brady, Alvarez, & Oliva, 2010).

The account outlined above maps neatly onto the findings in Chapters 3 and 4 of this dissertation: 24-month-olds spontaneously activate phonological representations in a free-viewing task, whereas adults in a very similar task do not. It also explains the discrepancy between the absence of phonological activation in Chapter 4 and the results of Chapter 2, where adults in a working memory task do activate phonological labels. In both cases, adults were arguably selectively recruiting the representations best suited for the task at hand. Decreasing automatic verbal encoding and the rise of strategic verbal encoding through development is consistent with the absence of implicit verbal activation in 4-year-olds in Chapter 2’s working memory study. On this story, phonological activation is no longer automatic in preschoolers and they have not yet acquired the metacognitive knowledge to realize when verbal encoding would be strategic to invoke. While this developmental account does explain the current data pattern, it also has unsatisfying aspects. Why is phonological activation automatic in infants? Does it serve a particular

cognitive function at this age? In addition, there are questions about the end of automatic implicit verbalization. Do adults inhibit verbal activation (e.g. would phonological activation be observed under cognitive load conditions?) or is there a developmental discontinuity such that this activation no longer occurs?

Another hypothesis is that, at all developmental stages, self-directed speech can occur at different levels of linguistic representation, from abstract implicit linguistic representations to overt private speech. The level at which activation occurs depends on the task at hand, the individual's familiarity with that task, and meta-cognitive factors. Of course, the latter two factors are both likely to correlate with age. The interaction of age and task in terms of verbal activation can be seen across the studies in this dissertation. For example, comparing the results from Chapter 4 to infant studies (e.g. in Chapter 3, and see also Mani et al., 2012; Mani & Plunkett, 2010; 2011), it seems that verbal activation during picture viewing occurs at a higher or more central level in adults than in infants. Although the results in Chapter 4 are open to a few different interpretations, one possible conclusion is that adults were activating lexeme-level representations while processing the visual images, whereas infants in a similar experimental paradigm were activating phoneme-level representations.

It is reasonable to imagine that the level of verbal representation that is functional to encode in a task changes with age. Phonological activation during picture viewing might be a useful activity for a 2-year-old who is less familiar with both the words and the concepts at hand. The notion that implicit verbal activation during picture viewing becomes less concrete (or less phonological) with age is indirectly supported by a study on overregularization by Ramscar, Dye and McCauley (in press). In this study, priming

younger children with images of sets of items, which corresponded to regular plurals, increased overregularization rates when children subsequently described images that should have evoked irregular plurals. On the other hand, the same primes decreased overregularization rates in older children. While the authors interpret these results in light of accumulating evidence about plural forms through language learning, this pattern is also consistent with an explanation where the younger children were more likely to activate the phonological form of the labels of the plural sets, and therefore the phonological form of the regular plural morpheme was more likely to interfere at the time of test. Of course, there are other possible mechanisms that could lead to the result, including changes in the underlying representations in the plural forms of various nouns, but it is interesting to consider the role implicit verbal activation might have been playing in this paradigm.

The idea here is not that implicit verbal representations are necessarily more concrete (e.g. at a phonological versus lexical level of representation) in younger populations, but that the level of representation depends on how the individual interacts with the task at hand. One specific suggestion is that more concrete representations are more useful when individuals are less familiar with the cognitive task at hand. The thought that implicit verbal representations become less concrete with increases in task proficiency and experience, which is of course related to age, echoes an idea put forward with respect to overt private speech. While the Vygotskian tradition tends to view the internalization of private speech as a developmental stage, there are reasons to believe that the shift is task-dependent. In particular, there is evidence that adults performing difficult or unfamiliar tasks sometimes revert to audible private speech, rather than strictly using silent inner speech (Sanchez Medina, Rubio, & De la Mata Benitez, 2009). Perhaps, rather than simply

observing a dichotomy between private speech and inner speech, it is more appropriate to think of a continuum of levels of representation of “inner” speech. On this continuum, the most concrete end of the scale would correspond to overt private speech, and the most abstract end would be approaching nonverbal conceptual representations.

Turning to larger issues, a more complete understanding implicit verbal activation has implications for many areas of cognitive psychology. A few examples are given here. First, accurate theories of the function of inner speech are important for understanding the consequences of atypical inner speech, such as in some subtypes of schizophrenia. There is also some evidence that children with autism spectrum disorders use inner speech differently (Whitehouse, Maybery, & Durkin, 2006). Second, studies on the nature and deployment of inner speech are critical to the interpretation of linguistic relativity effects. To the extent that a cognitive task leads to spontaneous verbal activation, then we should not be surprised to see differences in performance across language groups, since the task evoked differing linguistic representations. Third, this research has implications for the study of language processing. Many studies within psycholinguistics use the Visual World Paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), where participants listen to spoken utterances which correspond to a visual display, while their eye movements are recorded. If participants are likely to encode the visual stimuli at some level of verbal representation, independently of the auditory stimuli, this must be taken into account in interpreting their eye movement patterns. If we neglect the possibility of linguistic activation stemming from the visual input, we might draw false conclusions about the efficiency and speed of processing of the auditory input. Moreover, the study of implicit

verbal labeling, and of inner speech more generally, might also be valuable in informing theories of how the non-linguistic context interacts with language processing.

In sum, the studies presented in this dissertation demonstrate that inner speech is present in infants as young as 24-months-old, that implicit linguistic representations depend on the task at hand, and that this task-dependence interacts with age. Taken together, these studies also illustrate how eye-tracking can be used as a window into internal verbal representations in paradigms that are appropriate for a wide range of populations. Still, there are many unanswered questions for future research to address. Of particular interest is investigation of the function of internal verbal representations. When and why is it useful to recruit language for non-communicative cognitive tasks? And if it is not useful, then why does it occur? And, in either case, what are the mechanisms that result in inner speech? A full answer to these questions will allow us to better understand the patterns of implicit verbal activation observed in these studies and others, to explain when we should expect to see inner speech across different contexts, and to predict what form inner speech will take.



## References

- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6, 451-474.
- Alario, F., Costa, A., Ferreira, V. S., & Pickering, M. J. (2006). Architectures, representations and processes of language production. *Language And Cognitive Processes*, 21, 777-789.
- Allik, J. P., & Siegel, A. W. (1976). The use of the cumulative rehearsal strategy: A developmental study. *Journal of Experimental Child Psychology*, 21, 316-327.
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements : Evidence for Continuous Mapping Models, *Journal of Memory and Language*, 38, 419-439.
- Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology*, 42, 117-131.
- Altmann, G.T.M. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, 92, 129-170.
- Bååth, R. (2010). ChildFreq: An Online Tool to Explore Word Frequencies in Child Language. *LUCS Minor*, 16.
- Bach, M. J., & Underwood, B. J. (1970). Developmental changes in memory attributes. *Journal of Educational Psychology*, 61, 292-296.
- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, 111, 329-344.
- Bates D., & Maechler M. (2010). *lme4: Linear mixed-effects models using S4 classes* (R package version 0.999375-34). Retrieved from <http://CRAN.R-project.org/package=lme4>.

- Berk, L. E. (1986). Relationship of elementary school children's private speech to behavioral accompaniment to task, attention, and task performance. *Developmental Psychology*, 22, 671–680.
- Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17, 89–98.
- Boroditsky, L., Ham, W. & Ramscar, M. (2002). What is universal about event perception? Comparing English and Indonesian speakers. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*. Fairfax, VA.
- Brown, R. M. (1977). An Examination of Visual and Verbal Coding Processes in Preschool Children. *Child Development*, 48, 38.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177-208.
- Caramazza, A., Bi, Y., Costa, A., & Miozzo, M. (2004). What determines the speed of lexical access: homophone or specific-word frequency? A reply to Jescheniak et al. (2003). *Journal of experimental psychology. Learning, memory, and cognition*, 30, 278–82.
- Chomsky, N. (1979). Language and responsibility. Pantheon: New York.
- Chomsky, N. (2007). Of Minds and Language. *Biolinguistics*, 1, 9-27.
- Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology*, 5, 398–405.
- Cramer, P. (1975). Learning pictures and words: The role of implicit verbal labeling. *Journal of Experimental Child Psychology*, 489–501.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125–127.
- Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

- Duncan, R., & Tarulli, D. (2009). On the persistence of private speech: Empirical and theoretical considerations. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 176-187). New York, NY US: Cambridge University Press.
- Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, 22, 49–68.
- Fernyhough, C. (2009). Dialogic thinking. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 42-52). New York, NY US: Cambridge University Press.
- Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: associated ERPs and effects of lateralization. *Brain and Cognition*, 59, 145–58.
- Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist*, 34, 906-911.
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, 37, 283–299.
- Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Development*, 68, 39.
- Flavell, J. H., & Wong, A. A. (2009). Young children's knowledge about overt and covert private speech. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 143-149). New York, NY US: Cambridge University Press.
- Fodor, Jerry (1975) *The language of thought*. New York: Crowell.
- Ford, S., & Silber, K. P. (1994). Working memory in children: A developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology*, 12, 165–175.
- Gathercole, S. E., Adams, a M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & cognition*, 22, 201–7.

- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind*. Cambridge, MA: MIT Press.
- Hayes, D. S., & Schulze, S. A. (1977). Visual encoding in preschoolers' serial retention. *Child Development, 48*, 1066-1070.
- Heavey, C. L. & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition, 17*, 798-810
- Henry, L. A. (1991). Development of auditory memory span: The role of rehearsal. *British Journal of Developmental Psychology, 9*, 493-511.
- Henry, L. A., Turner, J. E., Smith, P. T., & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory, 8*, 1-17.
- Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London B, 302*, 325-340.
- Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology, 7*, 347-362.
- Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory & Cognition, 17*, 175-185.
- Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language, 38*, 644-661.
- Hulme, C. (1984). Developmental differences in the effect of acoustic similarity on memory span. *Developmental Psychology, 20*, 650-652.
- Hulme, C., Silvester, J., Smith, S., & Muir, C. (1986). The effects of word length on memory for pictures: Evidence for speech coding in young children. *Journal of Experimental Child Psychology, 41*, 61-75.
- Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word-length, and acoustic similarity on serial-recall. *Journal of Experimental Child Psychology, 47*, 72-87.

- Indefrey, P. P., & Levelt, W. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101-144.
- Jackendoff, R. (1996). How language helps us think. *Pragmatics and Cognition*, 4 (1), 1–35.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jescheniak, J.D., Opperman, F., Hanstch, A., Wagner, V., Madebach, A., & Schriefers, H. (2009). Do perceived context pictures automatically activate their phonological code. *Experimental Psychology*, 56, 56-65.
- Johnston, R. S., & Conning, A. (1990). The effects of overt and covert rehearsal on the emergence of the phonological similarity effect in 5-year-old children. *British Journal of Developmental Psychology*, 8, 411–418.
- Johnston, R. S., Johnson, C., & Gray, C. (1987). The emergence of the word length effect in young children: The effects of overt and covert rehearsal. *British Journal of Developmental Psychology*, 5, 243–248.
- Kelly, M.H., Springer, K., & Keil, F.C. (1990). The relationship between syllable number and visual complexity in the acquisition of word meanings. *Memory & Cognition*, 18, 528-536.
- Khan, M., Fitts, W.S., & Snedeker, J. (2012). *Do adults spontaneously recruit linguistic labels during visual object processing?* Manuscript in preparation.
- Khan, M., Geojo, A., Wang, S., & Snedeker, J. (2013). *Spontaneous verbal encoding in 24-month-old infants*. Manuscript submitted for publication.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision research*, 46, 1762–1776.
- Klinger, E. & Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagination, Cognition and Personality*, 7, 105–128.

- Konkle, T., Brady, T. F., Alvarez, G.A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of experimental psychology. General*, 139, 558–78.
- Kosslyn, S. M. (1976). Using imagery to retrieve semantic information: A developmental study. *Child Development*, 47, 434–444.
- Kosslyn, S. M. (1994). *Image and brain*. Cambridge: MIT Press.
- Lenth, R. V. (2006-9). *Java applets for power and sample size* [Computer software]. Retrieved February 6, 2013, from <http://www.stat.uiowa.edu/~rlenth/Power>.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Brain and Behavioral Sciences*, 22, 1–75.
- Locke, J.L., & Fehr, F.S. (1970). Young children's use of the speech code in a recall task. *Journal of Experimental Child Psychology*, 10, 367-373.
- Lupyan, G. (2008). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology. General*, 137, 348–69.
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 671–85.
- McLeod, P. & Posner, M.I. (1984). Privileged loops from percept to act. In H. Bouma and D. Bowhuis (Eds.), *Attention and Performance X* (pp. 55-66). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Manfra, L. (2009). Preschool children's awareness and theory of speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech , executive functioning, and the development of verbal self-regulation*. Cambridge University Press: New York.
- Manfra, L., & Winsler, A. (2006). Preschool children's awareness of private speech. *International Journal of Behavioral Development* (30), 537-549.

- Mani, N. (2010). When Cup primes Dog: Phono-semantic priming in the toddler lexicon. *Proceedings of the 34th Annual Boston University Conference on Language Development* (pp. 291–302). Boston: Cascadilla Press.
- Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, 66, 612–622.
- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological science*, 21, 908–913.
- Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, 121, 196–206.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–90.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Martínez Manrique, F., & Vicente, A. (2010). “What the ...!”: The role of inner speech in conscious thought. *Journal of Consciousness Studies*, 17(9), 141–167.
- Mead, G.H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
- Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, 35(3), 494–503.
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14, 710–716.
- Meyer, A.S., & Damian, M.F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, 35, 494-503.
- Miller, G.A. (1996). *The science of words*. New York: W.H. Freeman.
- Miller, P.H. (1993). *Theories of developmental psychology*. New York: W. H. Freeman.

- Morisson, F.J., & Haith, M.M. (1976). Recognition versus recall of briefly presented visual information: a developmental study. *Journal of Experimental Child Psychology*, 21, 191-200.
- Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 555-563.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 70-84.
- Nagy, W.E. & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.) *The nature of vocabulary acquisition*. Hillsdale, NJ: Erlbaum.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251-269.
- Navarette, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence of a cascade model of lexical access. *Journal of Memory and Language*, 53, 359-377.
- Noizet, G., & Pynte, J. (1976). Implicit labelling and readiness for pronunciation during the perceptual process. *Perception*, 5, 217-223.
- Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development*, 26, 818-830.
- Piaget, J. (1962). *The language and thought of the child* (M. Gabain, trans). Cleveland, OH: Meridian. (Original work published in 1923).
- Pinker, S. (1994). *The language instinct*. New York: HarperCollins.
- Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental: Human Learning and Memory*, 2, 509-522.



- Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81, 10–15.
- Pressley, M., Borkowski, J. G., and O'Sullivan, J. (1985). Children's metamemory and the teaching of memory strategies. In Forrest-Pressley, D. L., MacKinnon, G. E., and Waller, T.G. (Eds.), *Metacognition, Cognition, and Human Performance* (pp. 111-153). New York: Academic Press.
- Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory-strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, 23, 270-288.
- Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80, 1-24.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3–900051-07–0.  
Retrieved from <http://www.R-project.org>.
- Ramscar, M., Dye, M. & McCauley, S. (in press). Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwort's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33, 217-236.
- Sánchez Medina, J. A., Alarcón Rubio, D., & De la Mata Benítez, M. L. Private Speech beyond Childhood: Testing the Developmental Hypothesis. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 188-98). New York, NY US: Cambridge University Press.
- Schneider, W., & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 224–257). Cambridge, UK: Cambridge University Press.
- Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20*. Hillsdale, NJ: Erlbaum.
- Snedeker, J. (2009). Children's sentence processing. In E. Bavin (Ed.), *The Handbook of Child Language* (pp. 331-338). Cambridge: Cambridge University Press.

- Spelke, E. S. (2003). What makes humans smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind* (pp. 277-311). Cambridge, MA: MIT Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Telling, A. L. (2008). *Semantic and phonological context effects in visual search*. (Unpublished doctoral dissertation). University of Birmingham, Birmingham.
- Vicente, A., & Martinez Manrique, F. (2011). Inner Speech: Nature and Functions. *Philosophy Compass*, 6, 209–219.
- Vygotsky, L.S. (1987). Thinking and speech (N. Minick, Trans.). In R.W. Rieber & A.S. Carton (Eds.), *The collected works of L.S. Vygotsky: Vol. 1. Problems of general psychology* (pp. 37-285). New York: Plenum. (Original work published 1934)
- Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 47, 857–65.
- Whorf, B. L. (1956). Science and linguistics. In J. B. Carroll (Ed.), *Language, thought and reality: selected writings of Benjamin Lee Whorf*. Cambridge; MA: MIT Press.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 7780–7785.
- Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3-41). New York, NY US: Cambridge University Press.
- Winsler, A., & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development*, 74, 659–78.

- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology. Learning, memory, and cognition*, 32(1), 1–14.
- Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. Nelson & M. Luciana (Eds), *Handbook of Developmental Cognitive Neuroscience* (2nd Ed.) (pp. 553-574). Cambridge, MA: MIT Press.
- Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, 11, 125-131.

## Appendix A

Item	Word Length	Frequency (per million words)	Categorization Speed	Visual Complexity Rating
arm	short	120	690.2	2.1
bell	short	46	616.4	2.6
bread	short	207	584.6666667	1.9
bus	short	292	608.2	4.7
corn	short	100	553.7777778	3.8
ear	short	152	595	3.1
fork	short	89	539.6	1.7
key	short	43	638.7777778	2.3
kite	short	43	597.5555556	2.7
skunk	short	14	654.7	4.6
alligator	long	42	623.2222222	3.9
banana	long	51	478.1111111	1.3
butterfly	long	142	539.7	4.6
dinosaur	long	216	475.8888889	5.1
gorilla	long	5	615.1	3.5
helicopter	long	15	581.7	5
potato	long	54	554.8	1.3
strawberry	long	53	554.5555556	2.5
umbrella	long	62	498.3333333	3.7
watermelon	long	62	612.8	2.3

*List of items for recognition memory experiment conducted with 4-year-olds, 7-year-olds and adults (Chapter 2). Frequencies extracted from CHILDES transcripts with 4-year-olds, average categorization speed from the norming task and visual complexity ratings (1=very simple, 7=quite complex) from the Amazon Mechanical Turk survey are also included.*

## Appendix B

Target	Distractor	Related Prime (SR, VS)	Related Subprime (SR)	Unrelated Prime (SR, VS)
apple	pen	ball (1.9, 2.6)	banana (5.8)	cup (1.3, 1.1)
bread	phone	key (1.0, 1.2)	cake (5.9)	sheep (1.2, 2.1)
bus	tree	carrot (1.3, 1.7)	car (5.9)	duck (1.1, 1.7)
coat	foot	house (2.4, 1.1)	hat (5.2)	ball (1.1, 1.4)
cow	flower	hand (1.2, 1.4)	horse (5.6)	book (1.0, 1.6)
dog	box	cup (1.5, 1.2)	cat (5.7)	house (2.8, 2.0)
sock	plane	sheep (1.8, 2.1)	shoe (5.8)	key (1.0, 1.1)
table	brush	cheese (1.8, 1.2)	chair (5.7)	carrot (1.5, 1.1)
train	spoon	book (1.5, 1.2)	boat (5.4)	hand (1.2, 1.0)
window	bib	duck (1.0, 1.4)	door (5.1)	cheese (1.1, 1.2)

*List of items for spontaneous implicit labeling experiment with 24-month-olds (Chapter 3). SR indicates the average rating for semantic relatedness to the target and VS indicates the average rating for visual similarity to the target.*